

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 00:46:24 ; Search time 650 Seconds
(without alignments)
7124.984 Million cell updates/sec

Title: US-09-914-543-45

Perfect score: 960
Sequence: 1 atgagcaagaaaagtctgt.....atagacctttatttcttaa 960

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3163042 seqs, 2412103800 residues

Total number of hits satisfying chosen parameters: 6326094

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/prodata/1/pubna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/1/pubna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/1/pubna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/1/pubna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/1/pubna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/1/pubna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/1/pubna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/1/pubna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/1/pubna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/1/pubna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/1/pubna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/1/pubna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/1/pubna/US09_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/1/pubna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/1/pubna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/1/pubna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/prodata/1/pubna/US10C_NEW_PUB.seq.*
- 18: /cgn2_6/prodata/1/pubna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/prodata/1/pubna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	960	100.0	960	14	US-10-121-032-60
2	960	100.0	960	15	Sequence 60, Appl
3	958.4	99.8	1134	14	Sequence 60, Appl
4	536.8	55.9	978	15	Sequence 4, Appl
5	534	55.6	903	15	Sequence 56, Appl
6	84.2	8.8	825	16	US-10-228-063-58
7	39	4.1	585	13	US-10-369-493-26728
8	39	4.1	585	16	Sequence 26728, A
9	37.8	3.9	728	8	Sequence 197533,
10	37.8	3.9	728	13	Sequence 197533,
11	37.2	3.9	2338	9	US-10-027-632-197533
12	36.8	3.8	520	15	Sequence 611, Appl
13	36.8	3.8	520	15	Sequence 611, Appl
14	36.8	3.8	520	15	Sequence 1558, Ap
15	36.8	3.8	520	15	Sequence 144, Appl
16	36.8	3.8	520	15	Sequence 144, Appl
17	36.8	3.8	520	15	Sequence 144, Appl
18	36.8	3.8	520	15	Sequence 144, Appl
19	36.8	3.8	520	15	Sequence 144, Appl
20	36.8	3.8	520	15	Sequence 144, Appl
21	36.8	3.8	520	15	Sequence 144, Appl
22	36.8	3.8	520	15	Sequence 144, Appl
23	36.8	3.8	520	15	Sequence 144, Appl
24	36.8	3.8	520	15	Sequence 144, Appl
25	36.8	3.8	520	15	Sequence 144, Appl
26	36.8	3.8	520	15	Sequence 144, Appl
27	36.8	3.8	520	15	Sequence 144, Appl
28	36.8	3.8	520	15	Sequence 144, Appl
29	36.8	3.8	520	15	Sequence 144, Appl
30	36.8	3.8	520	15	Sequence 144, Appl
31	36.8	3.8	520	15	Sequence 144, Appl
32	36.8	3.8	520	15	Sequence 144, Appl
33	36.8	3.8	520	15	Sequence 144, Appl
34	36.8	3.8	520	15	Sequence 144, Appl
35	36.8	3.8	520	15	Sequence 144, Appl
36	36.8	3.8	520	15	Sequence 144, Appl
37	36.8	3.8	520	15	Sequence 144, Appl
38	36.8	3.8	520	15	Sequence 144, Appl
39	36.8	3.8	520	15	Sequence 144, Appl
40	36.8	3.8	520	15	Sequence 144, Appl
41	36.8	3.8	520	15	Sequence 144, Appl
42	36.8	3.8	520	15	Sequence 144, Appl
43	36.8	3.8	520	15	Sequence 144, Appl
44	36.8	3.8	520	15	Sequence 144, Appl
45	36.8	3.8	520	15	Sequence 144, Appl

ALIGNMENTS

RESULT 1
US-10-121-032-60
Sequence 60, Application US/10121032

Publication No. US2002015550A1

GENERAL INFORMATION:

APPLICANT: Bylina, Edward J.

TITLE OF INVENTION: GLYCOSIDASE ENZYMES

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Gray Cary Ware & Freidenrich LLP

STREET: 4365 Executive Drive, Suite 1600

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/121,032

FILING DATE: 09-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/134,078

FILING DATE: 13-AUG-1998

APPLICATION NUMBER: 08/949,026

FILING DATE: 10-OCT-1997

APPLICATION NUMBER: 60/056,916

FILING DATE: 06-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09010/024002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 858/677-1456

TELEFAX: 858/677-1465

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 960 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 1...957

SEQUENCE DESCRIPTION: SEQ ID NO: 60:

US-10-121-032-60

Query Match 100.0%; Score 960; DB 14; Length 960;

Best Local Similarity 100.0%; Pred. No. 3.4e-272;

Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCAAGAAAAGTTGGTATCTATCTTAAACAATCTTTAGTACAGCAATA 60
 DB 1 ATGAGCAAGAAAAGTTGGTATCTATCTTAAACAATCTTTAGTACAGCAATA 60
 QY 61 TATTTGTGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
 DB 61 TATTTGTGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
 QY 121 CCACCCCAACACACTTCCACTACCAAGTTCTCAAGATTAGATACCTGATGACGGT 180
 DB 121 CCACCCCAACACACTTCCACTACCAAGTTCTCAAGATTAGATACCTGATGACGGT 180
 QY 181 GAGTGCCAGGAGCTCTTATGATAGGATGTTGAGGAAACCCAGAAATCTACATTCAA 240
 DB 181 GAGTGCCAGGAGCTCTTATGATAGGATGTTGAGGAAACCCAGAAATCTACATTCAA 240
 QY 241 ATAAACCTATGAAACATTTCTTAATGCTACTGATGATGCTGATGATGATGATGAT 300
 DB 241 ATAAACCTATGAAACATTTCTTAATGCTACTGATGATGCTGATGATGATGATGAT 300
 QY 301 AGCGGCGTCTTCACTAGTCCCAACAACTTGACAACTTGTCTTGAGGATAGAGTAAT 360
 DB 301 AGCGGCGTCTTCACTAGTCCCAACAACTTGACAACTTGTCTTGAGGATAGAGTAAT 360
 QY 361 TGGGTGCAATGATACCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
 DB 361 TGGGTGCAATGATACCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
 QY 421 ACTGATGCCCCAATACCAATACCCAGTAAAGTTTCAACCTTAACAGACTTCTATCTAACA 480
 DB 421 ACTGATGCCCCAATACCAATACCCAGTAAAGTTTCAACCTTAACAGACTTCTATCTAACA 480
 QY 481 ATCTCTATAAATGAGCCCAAGAACCGCTGCGCAATTAATCTCGCAATAGAAATCTCGG 540
 DB 481 ATCTCTATAAATGAGCCCAAGAACCGCTGCGCAATTAATCTCGCAATAGAAATCTCGG 540
 QY 541 TTAACGAGAGAGCTTGAGAACACAGGATTAACAGCGATGAGCAAGAGTAATGATA 600
 DB 541 TTAACGAGAGAGCTTGAGAACACAGGATTAACAGCGATGAGCAAGAGTAATGATA 600
 QY 601 TGGATTTACTATGACGATTAACACCGCTGGCTCCCAAGTTAAGGAGATTTGATGCCA 660
 DB 601 TGGATTTACTATGACGATTAACACCGCTGGCTCCCAAGTTAAGGAGATTTGATGCCA 660
 QY 661 ATATAGTTACCGAAACACAGTAAATGCTACATTTGAAGTATGGAAGGCAACATTTGGT 720
 DB 661 ATATAGTTACCGAAACACAGTAAATGCTACATTTGAAGTATGGAAGGCAACATTTGGT 720
 QY 721 TGGGAGTATGTTGCAATTAAGAAAGACCCCAATCAAGAGGGAACAGTGCAATTTCCA 780
 DB 721 TGGGAGTATGTTGCAATTAAGAAAGACCCCAATCAAGAGGGAACAGTGCAATTTCCA 780
 QY 781 TACGAGCATTTAAGTGTGACGCCAACAATTTCAAGCTTACCAAAATTACACAGAACTT 840
 DB 781 TACGAGCATTTAAGTGTGACGCCAACAATTTCAAGCTTACCAAAATTACACAGAACTT 840

QY 841 TACTTAGAGGACGTGGAGATTGGAACCTGAGTTTGGAGCCAGCACTACTCGGCCAC 900
 DB 841 TACTTAGAGGACGTGGAGATTGGAACCTGAGTTTGGAGCCAGCACTACTCGGCCAC 900
 QY 901 CTAGAGTGGTGGATCACAACATACAACTAACTCTCTAGATAGACCTCTTATTTCTAA 960
 DB 901 CTAGAGTGGTGGATCACAACATACAACTAACTCTCTAGATAGACCTCTTATTTCTAA 960

RESULT 2

US-10-093-037-60

; Sequence 60, Application US/10093037

; Publication No. US2003078397A1

; GENERAL INFORMATION:

; APPLICANT: Jay M. Short

; APPLICANT: Bylina, Edward

; APPLICANT: Swanson, Ronald V.

; APPLICANT: Mathur, Eric J.

; APPLICANT: Lam, David E.

; TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF

; FILE REFERENCE: 09010-024006

; CURRENT APPLICATION NUMBER: US/10/093, 037

; PRIOR FILING DATE: 2002-03-06

; PRIOR APPLICATION NUMBER: US 09/910, 579

; PRIOR FILING DATE: 2001-07-20

; PRIOR APPLICATION NUMBER: US 09/134, 078

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 08/949, 026

; PRIOR FILING DATE: 1997-10-10

; PRIOR APPLICATION NUMBER: US 60/056, 916

; PRIOR FILING DATE: 1996-12-06

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 60

; LENGTH: 960

; TYPE: DNA

; ORGANISM: Pyrococcus furiosus

US-10-093-037-60

Query Match

Best Local Similarity 100.0%; Score 960; DB 15; Length 960;

Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCAAGAAAAGTTGGTATCTATCTTAAACAATCTTTAGTACAGCAATA 60
 DB 1 ATGAGCAAGAAAAGTTGGTATCTATCTTAAACAATCTTTAGTACAGCAATA 60
 QY 61 TATTTGTGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
 DB 61 TATTTGTGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
 QY 121 CCACCCCAACACACTTCCACTACCAAGTTCTCAAGATTAGATACCTGATGACGGT 180
 DB 121 CCACCCCAACACACTTCCACTACCAAGTTCTCAAGATTAGATACCTGATGACGGT 180
 QY 181 GAGTGCCAGGAGCTCTTATGATAGGATGTTGAGGAAACCCAGAAATCTACATTCAA 240
 DB 181 GAGTGCCAGGAGCTCTTATGATAGGATGTTGAGGAAACCCAGAAATCTACATTCAA 240
 QY 241 ATAAACCTATGAAACATTTCTTAATGCTACTGATGATGCTGAGATGACGTAACCTTAA 300
 DB 241 ATAAACCTATGAAACATTTCTTAATGCTACTGATGATGCTGAGATGACGTAACCTTAA 300
 QY 301 AGCGGCGTCTTCACTAGTCCCAACAACTTGACAACTTGTCTTGAGGATAGAGTAAT 360
 DB 301 AGCGGCGTCTTCACTAGTCCCAACAACTTGACAACTTGTCTTGAGGATAGAGTAAT 360
 QY 361 TGGGTGCAATGATACCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
 DB 361 TGGGTGCAATGATACCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
 QY 421 ACTGATGCCCCAATACCAATACCCAGTAAAGTTTCAACCTTAACAGACTTCTATCTAACA 480

Db 421 ACTGATGGCCCAATACCATTTACCCAGTAAGTTTCACACCTTAACAGACTTCTATCTAACA 480
 Qy 481 ATCTCCTATTAACCTTGAGCCCAAGAACGGGCTGCGCAATTAACCTTCGCAATAGAAATCCTGG 540
 Db 481 ATCTCCTATTAACCTTGAGCCCAAGAACGGGCTGCGCAATTAACCTTCGCAATAGAAATCCTGG 540
 Qy 541 TTAACGAGAGAGCTTGGAGAACCAAGAGAAATTAACAGCGATGAGCAAGAAAGTAATGATA 600
 Db 541 TTAACGAGAGAGCTTGGAGAACCAAGAGAAATTAACAGCGATGAGCAAGAAAGTAATGATA 600
 Qy 601 TGGATTTACTATGACGATTAACACCGGCTGCTGCCAAAGTTAAGGAGATTTAGTCCCA 660
 Db 601 TGGATTTACTATGACGATTAACACCGGCTGCTGCCAAAGTTAAGGAGATTTAGTCCCA 660
 Qy 661 ATAATAGTTAAGCAAGACCAAGTAATGCTACATTTGCAAGTATGGAAGGCAAAATTTGGT 720
 Db 661 ATAATAGTTAAGCAAGACCAAGTAATGCTACATTTGCAAGTATGGAAGGCAAAATTTGGT 720
 Qy 721 TGGAGATATGTTGCAATTTAGAAATAAGACCCCAATCAAGAGGGGAAACAGTGACAAATTTCCA 780
 Db 721 TGGAGATATGTTGCAATTTAGAAATAAGACCCCAATCAAGAGGGGAAACAGTGACAAATTTCCA 780
 Qy 781 TACGAGCATTTATAAGTTTGGAGCCCAACATTTTCAGCTTACCAAAATTTACACAGAACTT 840
 Db 781 TACGAGCATTTATAAGTTTGGAGCCCAACATTTTCAGCTTACCAAAATTTACACAGAACTT 840
 Qy 841 TACTTAGAGAGCTGGAGATTTGGAATGGAATGAGTTTGGAAAGCCCAAGCACTACCTCGGCCAC 900
 Db 841 TACTTAGAGAGCTGGAGATTTGGAATGGAATGAGTTTGGAAAGCCCAAGCACTACCTCGGCCAC 900
 Qy 901 CTAGAGTGGTGATGACAAAACATTAACCTAACTCTCTAGATAGACCTCTTATTTCTTAA 960
 Db 901 CTAGAGTGGTGATGACAAAACATTAACCTAACTCTCTAGATAGACCTCTTATTTCTTAA 960

RESULT 3

US-10-003-759-4
 ; Sequence 4, Application US/10003759
 ; Publication No. US20020102699A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Krivysztov B.
 ; APPLICANT: Holst, Olof Peder
 ; APPLICANT: Hachem, Maher Youssef Abou
 ; APPLICANT: Karlsson, Eva Margareta No. US20020102699A1dberg
 ; APPLICANT: Hreggvidsson, Gudmundur O.
 ; TITLE OF INVENTION: Thermostable Cellulase
 ; FILE REFERENCE: P5099PC00
 ; CURRENT APPLICATION NUMBER: US/10/003,759
 ; CURRENT FILING DATE: 2001-10-23
 ; PRIOR APPLICATION NUMBER: PCT/IS01/00012
 ; PRIOR FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: 09/594,884
 ; PRIOR FILING DATE: 2000-06-15
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 1134
 ; TYPE: DNA
 ; ORGANISM: Pyrococcus furiosus
 ; FEATURE:
 ; NAME/KEY: exon
 ; LOCATION: (82)..(1041)
 ; OTHER INFORMATION:
 ; US-10-003-759-4

Query Match 99.8%; Score 958.4; DB 14; Length 1134;
 Best Local Similarity 99.9%; Pred. No. 1.1e-271;
 Matches 959; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 ATGACGAGAAAGTTCGTATCGTACTATCTTAAACATCTTTAGTACAGCAATA 60
 Db 82 ATGACGAGAAAGTTCGTATCGTACTATCTTAAACATCTTTAGTACAGCAATA 141

Qy 61 TATTTTGTAGAAAGTATCATACCTCTGAGGCAAGTCAACCTCAAAATACCTCATCTACA 120
 Db 142 TATTTTGTAGAAAGTATCATACCTCTGAGGCAAGTCAACCTCAAAATACCTCATCTACA 201
 Qy 121 CCACCCCAACAACTTTCCTACTACCAAGGTTCTCAAGATTAGATACCTCATACCGGT 180
 Db 202 CCACCCCAACAACTTTCCTACTACCAAGGTTCTCAAGATTAGATACCTCATACCGGT 261
 Qy 181 GAGTGGCCAGGAGCTTCTATTGATAAGGATGCTGATGGGAAACCCAGAAATTTCTACATTTGAA 240
 Db 262 GAGTGGCCAGGAGCTTCTATTGATAAGGATGCTGATGGGAAACCCAGAAATTTCTACATTTGAA 321
 Qy 241 ATAAACCTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTACAAATTTAAC 300
 Db 322 ATAAACCTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTACAAATTTAAC 381
 Qy 301 AGCGGCTCTTCACTACGTCCAAACATTTGCAACATTTCTTCTGAGGATAGAAATTAAT 360
 Db 382 AGCGGCTCTTCACTACGTCCAAACATTTGCAACATTTCTTCTGAGGATAGAAATTAAT 441
 Qy 361 TGGGTGCATGGATACCCCGAAATATTTCTATGAAAACAAAGCCATGGAATGCAAACTACGCA 420
 Db 442 TGGGTGCATGGATACCCCGAAATATTTCTATGAAAACAAAGCCATGGAATGCAAACTACGCA 501
 Qy 421 ACTGATGGCCCAATACCATTTCCAGTAAAGTTTCAAACTTAACAGACTTCTATCTAACA 480
 Db 502 ACTGATGGCCCAATACCATTTCCAGTAAAGTTTCAAACTTAACAGACTTCTATCTAACA 561
 Qy 481 ATCTCTCTATAAACTTGGAGCCCAAGACGGCTGCTCCAAAGTTAAAGAGATTGTTAGTCCG 540
 Db 562 ATCTCTCTATAAACTTGGAGCCCAAGACGGCTGCTCCAAAGTTAAAGAGATTGTTAGTCCG 621
 Qy 541 TTAACGAGAGAGCTTGGAGAACCAAGAAATTAACAGCGATGAGCAAGAAATTAATGATA 600
 Db 622 TTAACGAGAGAGCTTGGAGAACCAAGAAATTAACAGCGATGAGCAAGAAATTAATGATA 681
 Qy 601 TGGATTTTACTATGACGATTTACAAACCGCTGCTCCAAAGTTAAAGAGATTGTTAGTCCCA 660
 Db 682 TGGATTTTACTATGACGATTTACAAACCGCTGCTCCAAAGTTAAAGAGATTGTTAGTCCCA 741
 Qy 661 ATAATAGTTAAGCAACACAGTAAATGCTACATTTGAAAGTATGGAAGGCAAAATTTGGT 720
 Db 742 ATAATAGTTAAGCAACACAGTAAATGCTACATTTGAAAGTATGGAAGGCAAAATTTGGT 801
 Qy 721 TGGAGTATGTTGCTATTTAGAAATAAGACCCCAATCAAGAGGGGAAACAGTGACAAATTTCCA 780
 Db 802 TGGAGTATGTTGCTATTTAGAAATAAGACCCCAATCAAGAGGGGAAACAGTGACAAATTTCCA 861
 Qy 781 TACGAGCATTTTATAAGTGTGTCAGCCCAACATTTCAAGCTTACCAAAATTTACAGAACTT 840
 Db 862 TACGAGCATTTTATAAGTGTGTCAGCCCAACATTTCAAGCTTACCAAAATTTACAGAACTT 921
 Qy 841 TACTTAGAGAGCTGGAGATTTGAACTGATTTGGAAAGCCCAAGCACTACCTCGGCCAC 900
 Db 922 TACTTAGAGAGCTGGAGATTTGAACTGATTTGGAAAGCCCAAGCACTACCTCGGCCAC 981
 Qy 901 CTAGAGTGGTGATGACAAAACATTAACCTAACTCTCTAGATAGACCTCTTATTTCTTAA 960
 Db 982 CTAGAGTGGTGATGACAAAACATTAACCTAACTCTCTAGATAGACCTCTTATTTCTTAA 1041

RESULT 4

US-10-228-063-58
 ; Sequence 58, Application US/10228063
 ; Publication No. US20030135885A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lananhan, Mike
 ; TITLE OF INVENTION: Self-processing Plants and Plant Parts
 ; FILE REFERENCE: 109846.317
 ; CURRENT APPLICATION NUMBER: US/10/228,063
 ; CURRENT FILING DATE: 2002-12-12
 ; NUMBER OF SEQ ID NOS: 60

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 978
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-228-063-58

Query Match          55.9%; Score 536.8; DB 15; Length 978;
Best Local Similarity 74.4%; Pred. No. 2.3e-147;
Matches 676; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

QY 52 CAGGCAATATATTTGTAGAAAGATCATACCTCTGAGGACAACTCAACTTCAATACC 111
Db 70 CATCAATCTACTTGTGAGAGTACCACACTCCGAGGACAACTCCACCTCCAACACC 129
QY 112 TCATCTACACACCCCAACACACTTTCCACTACCAAGTTTCTCAAGATTAGATACCT 171
Db 130 TCTCCACCCCGCGAGACACCTCTCCACCAACAGGTGCTCAAGATCGCTACCG 189
QY 172 GATGACGGTGAAGTGGCCAGGAGCTCTATTGATAGAGTGGTGGAAACCCAGAAATTC 231
Db 190 GACGACGGTGAAGTGGCCCGCCCGGATCGACAAAGGACGGCGCAACCCCGAGTTC 249
QY 232 TACATTGAAATAAACCCTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTAC 291
Db 250 TACATCGAGATCAACTCTGAGACATCTCAACGCCACCGCTTCGCCGAGATGACCTAC 309
QY 292 AATTAAACAGCGGGCTCTTCACTAGCTCCAACTTGACAACTTGACAACTTCTTGAGGAT 351
Db 310 AACCTCACTAGTGGGTGCTCCACTACGTGAGCAGCTCGACAACTCGTCTCGCGAC 369
QY 352 AGAAGTAAATGGTGCATGATACCCCGAATATCTATGGAACAGACCCATGAAATGCA 411
Db 370 CCGTCAACTGGGTGACGGTACCCGGAATCTTCTACGGCAACAGCCGTGGAACGCC 429
QY 412 AACTACGCAACTGATGGCCCAATACCATTAACCAAGTAAAGTTTCAAACTAACAGACTTC 471
Db 430 AACTACGCAACAGCGGCCGATCCGCTCCGCTCCAAAGGTGTCACACTCACCGACTTC 489
QY 472 TATCTAACTATCTCTATAAATTTAGCGCCAGAACGSCCTGCCAATTAATTTGCAATA 531
Db 490 TACCTCAACATCTCTACAAGCTGAGCGGAGAACGGTCTCCCGATCAACTTCGCCATC 549
QY 532 GAATCCTGGTTAACGAGAAAGCTTTGGAGAACACAGGAAATTAACAGCGATGAGCAAGAA 591
Db 550 GAGTCTGGCTCACCCGCGAGGCTGGCGACCCAGCCGATCACTCCGACGAGCAGGAG 609
QY 592 GTAATGATATGATTTACTATGACGGATTAACAACGGCTGGCTCCAAAGTTAAGGATTC 651
Db 610 GTGATGATCTGATCTACTACGACGGCTCCAGCCCGCGGCTCCAAAGGTGAAGGATC 669
QY 652 GTAGTCCCAATTAATAGTTAAACGGAACACAGTAATGCTACATTTGAAGTATGAGGCA 711
Db 670 GTGGTCCGATCATGCTGAAACGGACCCCGGTGACGCCACTTCGAGGTGGAAGGCC 729
QY 712 AACATTGGTGGAGTATGTTGCAATTTAGATAATAAGACCCCAATCAAGAGGGGAACAGTG 771
Db 730 AACATCGGTGGAGTACGTGGCTTCGCGATCAAGACCCCGATCAAGAGGGGACCGTG 789
QY 772 ACAATTTCATCGGAGCATTTATAGTTTCCAGCCACATTTCAAGCTTACCAATATAC 831
Db 790 ACCATCCGTTAGGGCCCTTCACTCCGGTGGCCGCAACATCTCTCTCCCGAATAC 849
QY 832 ACAGAACTTTACTTAGAGGACGTGGAGATTGGAACTGAGTTTGGAAACGCCAACACTACC 891
Db 850 ACCGAGAAAGTACCTCGAGACGTGGAGATGGCAGCCGAGTTCCGACACCCCGTCCACCA 909
QY 892 TCCGCCCACTAGAGTGGTGGATCACAACATACATTAATCTCTCTAGATAGACCTTT 951
Db 910 TCCGCCCACTCGAGTGGTGGATCACAACATACATCACTCCCTACCCCGCTCGAGCCCGCTC 969
```

```
QY 952 ATTTCTCTA 959
Db 970 ATCTCTA 977

RESULT 5
US-10-228-063-56
; Sequence 56, Application US/10228063
; Publication No. US20030135885A1
; GENERAL INFORMATION:
; APPLICANT: Lanahan, Mike
; TITLE OF INVENTION: Self-processing Plants and Plant Parts
; FILE REFERENCE: 109846 317
; CURRENT APPLICATION NUMBER: US/10/228, 063
; CURRENT FILING DATE: 2002-12-12
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
; US-10-228-063-56

Query Match          55.6%; Score 534; DB 15; Length 903;
Best Local Similarity 74.5%; Pred. No. 1.5e-146;
Matches 672; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 58 ATATATTTGTAGAAAGTATCATACCTCTGAGGACAACTCAACTTCAATACCTCATCT 117
Db 1 ATCTACTTCTGTGAGAGTACCACACTCCGAGGACAACTCCACCTCCAACACTCTCTCC 60
QY 118 ACACCAACCCCAACAACTTTCCACTACCAAGTTTCTCAAGATTAGATACCTCTGATGAC 177
Db 61 ACCCCGCGGACACCACTCTCCACCAACAGGTGCTCAAGATCGGTACCCGAGCAGAC 120
QY 178 GGTGAGTGGCCAGGAGCTCTTATGATAAGGATGGTGAATGGGAACCCAGAAATTCACATT 237
Db 121 GGCGAGTGGCGGGCGCCCGATCGACAGGACGGCGACCGCAACCCGAGTTCTACATC 180
QY 238 GAAATAAACCCTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACGATACAAATTA 297
Db 181 GAGATCAACTCTGGAACATCTCTAACGCCACCGCTTCGCCGAGATGACCTTACAACTTC 240
QY 298 ACACGCGGCTCTTCACTACGTCCAAACAACTTGACAACTTGTCTTGAGGATGAGAAT 357
Db 241 ACTAGTGGGTGCTCTCACTACGTGACGAGCTCGACAACTCGTCTCGCGACCGCTCC 300
QY 358 AATTGGTGCATGGATACCCCGAATATCTATGGAACAGCCATGGAATGCAACTAC 417
Db 301 AACTGGGTGACGGCTTACCCGGAATCTTCTACGGCAACAGCCGTGGAACGCCAATAC 360
QY 418 GCAACTGATGGCCCAATACCATTAACCAAGTTTCAAACTAACAGACTTCTATCTA 477
Db 361 GCCACGACGGCCCGATCCCGCTCCCGTCCAAAGGTGTCACACTCACCGACTTCTACCTC 420
QY 478 ACATCTCTATAAATTTGAGCCCAAGAACGSCCTGCCAATTAATTTGCAATAGAAATCC 537
Db 421 ACCATCTCTACAAGCTCGAGCCGAGAACGGTCTCCCGATCAACTTCGCCATCCAGTCC 480
QY 538 TGGTTAAACGAGAGAGCTTTGGAGAACACAGGAAATTAACAGCGATGAGCAAGAACTAATG 597
Db 481 TGGCTCACCCGAGGGCTGGCGACACCCGCACTCACTCCGACGACGAGGATGATG 540
QY 598 ATATGATTTACTATGACGGATTAACAACGGGTGGCTCCAAAGTTAAGGATTTAGTATGCT 657
Db 541 ATCTGATCTACTACGACGGCTCCAGCCCGCGGCTCCAAAGGTGAGGAGATCTGTGTTG 600
QY 658 CCAATAATAGTTAAACGGAACACAGTAAATGCTACATTTGAAGTATGGAAGCAAACTT 717
Db 601 CGATCATCTGTAAGCGGACACCCCGGTGAACGCCACTTCGAGGTGTGAAGGCCAATC 660
QY 718 GGTGGAGTATGTTTGCATTTTAGAATAAAGACCCCAATCAAGAGGGGAACAGTACAAT 777
```


Db 661 GGCTGGAGTAGTCTGGGCTTCCGCATCAAGACCCCGATCAAGAGGGGCGCCGTCACCATC 720
 QY 778 CCATACGGAGCATTTATAGTGTTCGAGCCCAACATTTCAAGCTTACCAATTTACACAGAA 837
 Db 721 CCGTACGGCGCTTCACTCTCGTGGCCGCAACATCTCTCCCTCCGAACTACACCGAG 780
 QY 838 CTTTACTTTAGAGGAGTGGAGATGGAACTGAGTTTGAACGCGCAAGCACTACCTCCGCC 897
 Db 781 AAGTACCTCGAGGAGCTGGAGATCGGACCGAGTGGGACCCCGTCCACCACTCCGCC 840
 QY 898 CACCTAGAGTGGTGGATCACAACATACATACATCTCTCTAGATAGCTCTTATTTCC 957
 Db 841 CACCTCGAGTGGTGGATCACAACATACATCTCTCTAGATAGCTCTTATTTCC 900
 QY 958 TA 959
 Db 901 TA 902

RESULT 6
 US-10-369-26728
 ; Sequence 26728, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 26728
 ; LENGTH: 825
 ; TYPE: DNA
 ; ORGANISM: Thermotoga maritima
 US-10-369-26728

Query Match 8.8%; Score 84.2; DB 16; Length 825;
 Best Local Similarity 50.2%; Pred. No. 6, 7e-14;
 Matches 305; Conservative 0; Mismatches 278; Indels 24; Gaps 3;

QY 332 ACAACATTGTCTTGGAGATGAAAGTAAATGGTGCATGGATACCCCGAAATATCTATG 391
 Db 206 ACAACATCGTTCTTCAAGATCCAGACAGCTGGTGCATGGATATCCAGAGATCTACTACG 265
 QY 392 GAAACAAGCCATGGATGCAACTAGCAACTGATGGCCCAATACCATTCACAGTAAAG 451
 Db 266 GTTACAAGCCCTGGCGGGGACAAACAGCGGTGTGAAATTTCTTCCTGTG-----AAGG 319
 QY 452 TTCAAACCTAACAGACTTCTATCTAACTCTCTATAAATCTTGAAGCCCAAGACCGCC 511
 Db 320 TGAAGATCTTCGGATTTCTACGTGACTCTTGTATCTCGATCTGGTACGAAACAACC 379
 QY 512 TCCCAATTAATTCGCAATAGATCTCTGTTAAACAGAGAGCTTGGAGAACACAGAA 571
 Db 380 TGCCTATCAACCTTCCCATGGAGCTGGATCAAGAGTCCCGACCAAGCTCTG--- 436
 QY 572 TTAAACAGCATGAGCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 631
 Db 437 TTTCTTCGGGTGACCGGAGATCATGGTTGGTCTCAACAACAGTCTGTATGTCGCGCG 496
 QY 632 GCTCCAAAGTTAAGAGATTTGATGTCCTCAATTAATAGTTAAACGGAACACCAATAGCTA 691
 Db 497 GTCAGAAAGTGACGAGTTCTACTACACAGTGCAGATTAACGGCGGTGAAGCAGGAACAA 556
 QY 692 CATTGTAAGTAGTGAAGCAACATCTGTTGGAGTATGTTGATTTAGATAAAGACC 751

Db 557 AGTGGGATGTTTACTTTTGCCCGTGGGATGGGATTAACCTTCTTTAGACTGACAACAC 616
 QY 752 CAATCAAGAGGGAACAGTGAACAATTCATACGGAGCAATTTATAAGTGTTCAGGCCAACA 811
 Db 617 CGATGAAGAGGAAGGTGAANAATCAACGTGAAGCACTTCGTTAGAAAGCCGCGGAG 676
 QY 812 TT-----TCAGCTTACCAANTTACACGAACCTTTTACTTTAGAGGAGCTGG 856
 Db 677 TTGTCAGAAGCACTCAACGAGATAGACAATTTGAGAGAGCTGTATTTCTCGCTGGG 736
 QY 857 AGATTGGAAGTGGTTGGAACGCAAGCACTACCTCCGCCACCTTAGAGTGTGTGATCA 916
 Db 737 AGATCGAAGCGAGTTTGGAGATCCGACACACAGCGCGGAATTCGGCTGGACTTTCA 796
 QY 917 CRAACAT 923
 Db 797 GGGACTT 803

RESULT 7
 US-10-027-632-197533/c
 ; Sequence 197533, Application US/10027632
 ; Publication No. US20020198371A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 197533
 ; LENGTH: 585
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-027-632-197533

Query Match 4.1%; Score 39; DB 13; Length 585;
 Best Local Similarity 50.3%; Pred. No. 1.2; Indels 0; Gaps 0;
 Matches 93; Conservative 1; Mismatches 91;

QY 638 AAGTTAAGGAGATTTAGTCCCAATTAATAGTTAAACGGAACACCAAGTAAATGCTTACATTG 697
 Db 523 AAGTACACAGATATGCTAAGAAAGGAGAGAAATGGAATCATATAAATGCTTAAATTA 464
 QY 698 AAGTATGGAAGCAACATTTGTTGGAGTATGTTGCAATTTAGAAATAAGACCCCAATCA 757
 Db 463 AACTACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 404
 QY 758 AAGAGGGAACAGTGAACAATTTCCATACGAGCAATTTATAGTGTTCAGGCAACATTTCAA 817
 Db 403 AAGAGGAACAGTGAACAATTTCCATATTTATTTCCACTATATTTAGTAAATCACATTTAA 344
 QY 818 GCTTA 822
 Db 343 AAGTA 339

```

RESULT 8
US-10-027-632-197533/c
; Sequence 197533, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 197533
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-197533

Query Match 4.1%; Score 39; DB 16; Length 585;
Best Local Similarity 50.3%; Pred. No. 1.2;
Matches 93; Conservative 1; Mismatches 91; Indels 0; Gaps 0;

QY 638 AAGTTAAGGAGATGTTAGTCCCAATTAAGTTAAGCGAACACCGAGTAATGCTACATTG 697
Db 523 AAGTACACAGATATGCTTAAGAAAGAGAGAAATGGAATCATATAAATGCTAAATTAA 464
QY 698 AAGTATGGAAGCAACATTGTTGGAGTATGTTGCTATTTAGATAAAGACCCCAATCA 757
Db 463 AACTACAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 404
QY 758 AAGAGGGAACAGTGAACATTCATCAGGAGCATTTATAGTGTTCGACCCCAACATTCA 817
Db 403 AAGAGAAACAGTGAACATGCTTATATTTCCAACTATATTAGTAAATCACTTTAA 344

QY 818 GCTTA 822
Db 343 AAGTA 339

```

```

RESULT 9
US-08-781-986A-611/c
; Sequence 611, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2

```

```

; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 611:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 728 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-611

Query Match 3.9%; Score 37.8; DB 8; Length 728;
Best Local Similarity 45.5%; Pred. No. 3;
Matches 135; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 541 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAAGAAATGATA 600
Db 456 TTAATACCAATGAATGATAATCTACTTACTAGTGAATGTAATGATAGTAA 397
QY 601 TGGATTACTATGACGGATTACAAACCGCTGGCTCCAAAGTTAAGGAGATTGAGTCCCA 660
Db 396 ATACCTTCTTGATAGTTTGTCTCAATTCGATGTAAATTAATTCATGTAATCCAA 337
QY 661 ATATAGTTAAACGAGACACAGTAATGCTACATTTGAAAGTATGGAAGCAACATTTGGT 720
Db 336 GCTACAGATAAGGAGAAATTAACCAATCCATTTGCAATGGAGATAATCCTAGTTTCT 277
QY 721 TGGGAGTATGTTTCATTTAGATAAGACCCCAATCAAGAGGAGGACAGTGAATTTCA 780
Db 276 TGAAGGTAGCTGGAATATATAATTTGAATCCATTTAAACAATAGCTTTAGAGGTCA 217
QY 781 TACGGAGCATTTATAAGTGTTCGAGCAACATTTTCAAGCTTTACCAATTTACACAGAA 837
Db 216 GTTATAAATAAAGTAATCGAATCGAATTAATTTCAACACAGGTAATAATCGTGAA 160

RESULT 10
US-10-329-624-611/c
; Sequence 611, Application US/10329624
; Publication No. US20040043037A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text

```

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/329,624
 FILING DATE: 27-Dec-2002
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/956,171
 FILING DATE: October 20, 1997
 APPLICATION NUMBER: 60/009,861
 FILING DATE: January 5, 1996
 APPLICATION NUMBER: 08/781,986
 FILING DATE: January 3, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Mark J. Hyman
 REGISTRATION NUMBER: 46,789
 REFERENCE/DOCKET NUMBER: PB248P1D1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (240) 314-1224
 TELEFAX: (301) 309-8439
 INFORMATION FOR SEQ ID NO: 611:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 728 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 611:
 US-10-329-624-611

Query Match 3.9%; Score 37.8; DB 13; Length 728;
 Best Local Similarity 45.5%; Pred. No. 3;
 Matches 135; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 541 TTAACGAGAGAGCTTGAGAACACACAGGAATTAACGCGATGAGCAAGTAATGATA 600
 DB 456 TTAATACCAATGAATGATATATCTACTTACTTAGTCAATGTAATGATAGTAA 397
 QY 601 TGGATTACTATGACGGATTACACCGGCTGGCTCCAAAGTTAAGGAGATTGATCCCA 660
 DB 396 ATAACCTTCCTGTAGTGTTCCTCAATTCGATGTAATTAATTAATCAATGTAATCCAA 337
 QY 661 ATAATAGTTACGAGACACAGTAATGCTACATTTGAGTATGAGGCGAACAATTGGT 720
 DB 336 GCTACAGATAAAGGAATAATCAACATCCATTTGCAATGAGATAATCCTGTTGTCT 277
 QY 721 TGGGAGTATGTTGCATTTAGATAAAGACCCCAATCAAGAGGGAACAGTACAAATCCA 780
 DB 276 TGAAGTAGACTGGAATATATAATTAATGATCCCAATTAACAAATAGCTGTTAGAAGTCA 217
 QY 781 TACGAGCATTTATAAGTGTGCGCCCAACATTTCAAGCTTACCAATTTACACAGAA 837
 DB 216 GTTATAAACTAAAGTAATCGAAGCATTAATTAATTAATCAACACAGGTAATAATGGTGAA 160

RESULT 11
 US-09-917-800A-1558/c
 Sequence 1558, Application US/09917800A
 Patent No. US20020119462A1
 GENERAL INFORMATION:
 APPLICANT: Mendrick, Donna
 APPLICANT: Porter, Mark
 APPLICANT: Johnson, Kory
 APPLICANT: Castle, Arthur
 APPLICANT: Elashoff, Michael
 APPLICANT: Gene Logic, Inc.
 TITLE OF INVENTION: Molecular Toxicology Modeling
 FILE REFERENCE: 44921-5038-US
 CURRENT APPLICATION NUMBER: US/09/917,800A
 CURRENT FILING DATE: 2001-07-31
 PRIOR APPLICATION NUMBER: US 60/222,040
 PRIOR FILING DATE: 2000-07-31
 PRIOR APPLICATION NUMBER: US 60/222,880
 PRIOR FILING DATE: 2000-11-02
 PRIOR APPLICATION NUMBER: US 60/290,029
 PRIOR FILING DATE: 2001-05-11
 PRIOR APPLICATION NUMBER: US 60/290,645

PRIOR FILING DATE: 2001-05-15
 PRIOR APPLICATION NUMBER: US 60/292,336
 PRIOR FILING DATE: 2001-05-22
 PRIOR APPLICATION NUMBER: US 60/295,798
 PRIOR FILING DATE: 2001-06-06
 PRIOR APPLICATION NUMBER: US 60/297,457
 PRIOR FILING DATE: 2001-06-13
 PRIOR APPLICATION NUMBER: US 60/298,884
 PRIOR FILING DATE: 2001-06-19
 PRIOR APPLICATION NUMBER: US 60/303,459
 PRIOR FILING DATE: 2001-07-09
 NUMBER OF SEQ ID NOS: 1740
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1558
 LENGTH: 2338
 TYPE: DNA
 ORGANISM: Rattus norvegicus
 FEATURE:
 OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_012674
 US-09-917-800A-1558

Query Match 3.9%; Score 37.2; DB 9; Length 2338;
 Best Local Similarity 49.0%; Pred. No. 9.1;
 Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 550 GRAGCTTGGAGAACACACAGGAATTAACGCGATGAGCAAGTAATGATATGGATTAC 609
 DB 1997 GAATGTGGAGGGGTACAGAGTACCCAAACCTTAACAGGAGTCTTAAAGTCAATTTTC 1938
 QY 610 TATGACGGATTACACCGGCTGGCTCCAAAGTTAAGGAGATTGATGCCCAATAATAGTT 669
 DB 1937 TTTAGGGGTATGGCTTTTCATATATTGGCAGTAACAGTGCATGGTGTACACTCATGAGT 1878
 QY 670 AACGGAACACCAAGTAATGCTACATTTGAGTATGAGGCGAACAACATTTGGTTGGGAGTAT 729
 DB 1877 ATATGACGACGATTAATGGACTCTTTGAGCTATTAAAAAATACATGAAGTTAGGAAGGT 1818
 QY 730 GTTGCAATTTAGATAAAGACCC 751
 DB 1817 GAACTTTAGAAACGGGACTC 1796

RESULT 12
 US-10-142-426-144
 Sequence 144, Application US/10142426
 Publication No. US20040048333A1
 GENERAL INFORMATION:
 APPLICANT: Baker, Kevin P.
 APPLICANT: Beresini, Maureen
 APPLICANT: DeForge, Laura
 APPLICANT: Desnoyers, Luc
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Sherwood, Steven
 APPLICANT: Smith, Victoria
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Watanabe, Colin K
 APPLICANT: Wood, William
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P330R1C224
 CURRENT APPLICATION NUMBER: US/10/142,426
 CURRENT FILING DATE: 2002-05-09
 Prior Application removed - See file Wrapper or Palm
 NUMBER OF SEQ ID NOS: 550
 SEQ ID NO 144
 LENGTH: 520

```
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-426-144

Query Match      3.8%; Score 36.8; DB 13; Length 520;
Best Local Similarity 10.2%; Pred. No. 4.9;
Matches 53; Conservative 143; Mismatches 322; Indels 0; Gaps 0;

QY 225 AGAATTCTACATTGAATAAAGCTTGAACAACTTCTTAATGCTACTGATTTGCTGAGAT 284
Db 1 MVARVGLLLRALQLLWHLDAQPAERGQELRKEAEFLKXYGLNEQVPKAPTSTRFS 60
QY 285 GACGTACAATTTAACACGGCGGCTCTTCACTACGTCCCAACAACTTGACAACATGCTT 344
Db 61 DAIRAFQWVSQLPVSGVLDRLATLRQMTPRCGVTDNTSYAAWAERISDLFAHRHTKMRK 120
QY 345 GAGGGATAGAAGTAATTGGTGCATGGATACCCCGAAATATCTATGGAACAAGCCCATG 404
Db 121 KRFAKQGNKWKYKHLNVLNWPHELPEPAVRGAVRAAFQOLWSNVSALEFWEAPATGPAD 180
QY 405 GAATGCAAACTAGCAACTGATGGCCCAATACCATTTACCCAGTAAGTTTCAAACTTAAC 464
Db 181 IRLTFQGDHNDGLGNAFDGPGGALAHAFLLPRGEAHFDQDERWSLSRRRGRNLFVLAH 240
QY 465 AGACTTCTATCAACTCTCTATAAAGTTCGAGCCCAAGACGGCTGCGCAATTAACCTT 524
Db 241 EIGHTLGLTHSPAPRALMAPYKRLGRDALLSWDDVLAVQSLYKPLGGSVAVQLPGKLF 300
QY 525 CGCAATAGAAATCCTGTTAAACAGAGAACTTGGAGAACACAGGAATTAACAGGATGA 584
Db 301 TDFETWDSYSPQRRPETQPKYCHSFDATVDRQQQLYIFKGSHEVAADGNVSEPR 644
QY 585 GCAAGAGTAATGATGATGATTTACTATGACGATTAACACCGCTGCTGCTCAAGTTAA 644
Db 361 PLQERWVGLPNIENAAVSLNDGDFYFKGRCWFRGPKVWGLPQLCRAGGLPRHPDA 420
QY 645 GGAGATTGTAGTCCCAATATAGTTTAAACGGAACACAGTAATGCTACTTTGAAGTATG 704
Db 421 ALFFPPLRLILFKGARYVVLARGGLQVEPYPRSLQDWGGIPEVSGALPRPDGSIIF 480
QY 705 GAAGGCAAACTGTTGGAGTATGTTGCTATTGAA 742
Db 481 RDDRWRDLQAKLQATTSGRWATELPMWGCWHANGSSA 518

RESULT 13
US-10-123-155-144
; Sequence 144, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 144
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-144

Query Match      3.8%; Score 36.8; DB 15; Length 520;
Best Local Similarity 10.2%; Pred. No. 4.9;
Matches 53; Conservative 143; Mismatches 322; Indels 0; Gaps 0;

QY 225 AGAATTCTACATTGAATAAAGCTTGAACAACTTCTTAATGCTACTGATTTGCTGAGAT 284
Db 1 MVARVGLLLRALQLLWHLDAQPAERGQELRKEAEFLKXYGLNEQVPKAPTSTRFS 60
QY 285 GACGTACAATTTAACACGGCGGCTCTTCACTACGTCCCAACAACTTGACAACATGCTT 344
Db 61 DAIRAFQWVSQLPVSGVLDRLATLRQMTPRCGVTDNTSYAAWAERISDLFAHRHTKMRK 120
QY 345 GAGGGATAGAAGTAATTGGTGCATGGATACCCCGAAATATCTATGGAACAAGCCCATG 404
Db 121 KRFAKQGNKWKYKHLNVLNWPHELPEPAVRGAVRAAFQOLWSNVSALEFWEAPATGPAD 180
QY 405 GAATGCAAACTAGCAACTGATGGCCCAATACCATTTACCCAGTAAGTTTCAAACTTAAC 464
Db 181 IRLTFQGDHNDGLGNAFDGPGGALAHAFLLPRGEAHFDQDERWSLSRRRGRNLFVLAH 240
QY 465 AGACTTCTATCAACTCTCTATAAAGTTCGAGCCCAAGACGGCTGCGCAATTAACCTT 524
Db 241 EIGHTLGLTHSPAPRALMAPYKRLGRDALLSWDDVLAVQSLYKPLGGSVAVQLPGKLF 300
QY 525 CGCAATAGAAATCCTGTTAAACAGAGAACTTGGAGAACACAGGAATTAACAGGATGA 584
Db 301 TDFETWDSYSPQRRPETQPKYCHSFDATVDRQQQLYIFKGSHEVAADGNVSEPR 644
QY 585 GCAAGAGTAATGATGATGATTTACTATGACGATTAACACCGCTGCTGCTCAAGTTAA 644
Db 361 PLQERWVGLPNIENAAVSLNDGDFYFKGRCWFRGPKVWGLPQLCRAGGLPRHPDA 420
QY 645 GGAGATTGTAGTCCCAATATAGTTTAAACGGAACACAGTAATGCTACTTTGAAGTATG 704
Db 421 ALFFPPLRLILFKGARYVVLARGGLQVEPYPRSLQDWGGIPEVSGALPRPDGSIIF 480
QY 705 GAAGGCAAACTGTTGGAGTATGTTGCTATTGAA 742
Db 481 RDDRWRDLQAKLQATTSGRWATELPMWGCWHANGSSA 518

RESULT 14
US-10-146-731-144
; Sequence 144, Application US/10146731
; Publication No. US20030129692A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC32
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
```

; CURRENT APPLICATION NUMBER: US/10/146,731
; CURRENT FILING DATE: 2002-05-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 144
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-146-731-144

Query Match 3.8%; Score 36.8; DB 15; Length 520;
Best Local Similarity 10.2%; Pred. No. 4.9; Indels 0; Gaps 0;
Matches 53; Conservative 143; Mismatches 322; Indels 0; Gaps 0;
QY 225 AGAATTCCTACATTAACCAATGGAACATCTTAATGCTACTGATTCGTGAGAT 284
DB 1 MVARVGLLLRALQLLLWGHLDQAQPAERGGQBLRKEAEAFLEKYGYLNEQVPKAPTSTRFS 60
QY 285 GACGTACAAATTAACCAAGCGGCTCTTCACTAGCTCCAAACATTCACACATTTGCTT 344
DB 61 DAIRAFQWVSQLPVSGVLDRLATLRQMTPRCGVTDNTSYAAWAERISDLFARHETKMRK 120
QY 345 GAGGATAGAAATTAATGGTGCATGATACCCGCAATATTTCTATGGAACAAGCCATG 404
DB 121 KRFAKQGNKWKYQHLVNLVWPEHLPEPAVRGAVRAAFQVLSVNSVALEFWEAPATGPAD 180
QY 405 GAATGCAAACTACGCAATGATGCGCCCAATACCATTCACCAAGTAAAGTTTCAACCTAAC 464
DB 181 IRLTFFQGDHNDGLGNAPDGGALAHAFPRRGEAHFPQDERWSLSRRRGRNLFVVLAH 240
QY 465 AGACTTCTATCAAACTCTCTATAAATGAGCCCAAGACCGGCTGCCCAATTAATT 524
DB 241 EIGHTLGLTHSPAPRALMAPYKRLGRDALLSWDDVLAVQSLYKPLGGSVAVOLPGKLF 300
QY 525 CGCAATAGAAATCTGTTTAAACGAGAGAGCTTGAGACAACACAGGAATTAACAGCCATGA 584
DB 301 TDFTWDSYSPQRRPETGPKYCHSSPDALTVDROOQLYIFKGSHEFWAADGNVSEPR 360
QY 585 GCAAGAAGTAATGATGATGATTTACTATGAGGATTACACCGGCTGCCCTCCAAAGTTAA 644
DB 361 PLQERWVGLPPIEAAAASVINDGDFYFFKGRGRCWRFRGPKVWGLPQLCRAGGLPRHPDA 420
QY 645 GGAGATTGTAGTCCCAATATAGTTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATG 704
DB 421 ALFPPRLRLILFKGARYVVLARGQLQVEPYPRSLQDWGGIPEEVSGALPRPDGSIIF 480
QY 705 GAAGCAAAACATTTGGTGGAGTATGTTGCAATTTAGAA 742
DB 481 RDRYWRLDQAKLQATTSGRWATELPMWGWCHANSOSA 518

RESULT 15

US-10-140-472-144

; Sequence 144, Application US/10140472

; Publication No. US2003013888A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 144
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-472-144

Query Match 3.8%; Score 36.8; DB 15; Length 520;
Best Local Similarity 10.2%; Pred. No. 4.9; Indels 0; Gaps 0;
Matches 53; Conservative 143; Mismatches 322; Indels 0; Gaps 0;
QY 225 AGAATTCCTACATTAACCAATGGAACATCTTAATGCTACTGATTCGTGAGAT 284
DB 1 MVARVGLLLRALQLLLWGHLDQAQPAERGGQBLRKEAEAFLEKYGYLNEQVPKAPTSTRFS 60
QY 285 GACGTACAAATTAACCAAGCGGCTCTTCACTAGCTCCAAACATTCACCAATTTGCTT 344
DB 61 DAIRAFQWVSQLPVSGVLDRLATLRQMTPRCGVTDNTSYAAWAERISDLFARHETKMRK 120
QY 345 GAGGATAGAAATTAATGGTGCATGATACCCGCAATATTTCTATGGAACAAGCCATG 404
DB 121 KRFAKQGNKWKYQHLVNLVWPEHLPEPAVRGAVRAAFQVLSVNSVALEFWEAPATGPAD 180
QY 405 GAATGCAAACTACGCAATGATGCGCCCAATACCATTCACCAAGTAAAGTTTCAACCTAAC 464
DB 181 IRLTFFQGDHNDGLGNAPDGGALAHAFPRRGEAHFPQDERWSLSRRRGRNLFVVLAH 240
QY 465 AGACTTCTATCAAACTCTCTATAAATGAGCCCAAGACCGGCTGCCCAATTAATT 524
DB 241 EIGHTLGLTHSPAPRALMAPYKRLGRDALLSWDDVLAVQSLYKPLGGSVAVOLPGKLF 300
QY 525 CGCAATAGAAATCTGTTTAAACGAGAGAGCTTGAGACAACACAGGAATTAACAGCCATGA 584
DB 301 TDFTWDSYSPQRRPETGPKYCHSSPDALTVDROOQLYIFKGSHEFWAADGNVSEPR 360
QY 585 GCAAGAAGTAATGATGATGATTTACTATGAGGATTACACCGGCTGCCCTCCAAAGTTAA 644
DB 361 PLQERWVGLPPIEAAAASVINDGDFYFFKGRGRCWRFRGPKVWGLPQLCRAGGLPRHPDA 420
QY 645 GGAGATTGTAGTCCCAATATAGTTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATG 704
DB 421 ALFPPRLRLILFKGARYVVLARGQLQVEPYPRSLQDWGGIPEEVSGALPRPDGSIIF 480
QY 705 GAAGCAAAACATTTGGTGGAGTATGTTGCAATTTAGAA 742
DB 481 RDRYWRLDQAKLQATTSGRWATELPMWGWCHANSOSA 518

Search completed: July 2, 2004, 02:27:28

Job time : 653 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2004, 21:58:24 ; Search time 102 Seconds
(without alignments)
5223.067 Million cell updates/sec

Title: US-09-914-543-45

Perfect score: 960

Sequence: 1 atgagcaagaaaagttcgt.....atagacctttattctctaa 960

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	960	100.0	960	4	US-09-134-078-60
2	114.2	11.9	867	3	US-08-995-280C-1
3	53.8	5.6	7218	1	US-08-232-463-14
4	37.8	3.9	728	4	US-08-956-171E-611
5	36	3.8	936	4	US-09-636-382A-16
6	34.8	3.6	1242	4	US-09-134-000C-3076
7	34.8	3.6	3146	4	US-09-620-312D-277
8	34.8	3.6	15016	4	US-09-601-198-60
9	34.2	3.6	1161	4	US-09-328-352-1845
10	33.8	3.5	1664976	4	US-08-916-421B-1
11	33.4	3.5	3543	3	US-09-224-024-27
12	33.4	3.5	3543	5	PCT-US94-07802-27
13	33.2	3.5	612	3	US-09-385-982-400
14	32.8	3.4	832	4	US-09-621-976-2813
15	32.8	3.4	1422	1	US-08-319-704-5
16	32.4	3.4	3735	4	US-09-543-681A-2102
17	32.4	3.4	90541	4	US-09-759-359A-3
18	32.2	3.4	645	4	US-09-328-352-661
19	32	3.3	1652	4	US-09-375-140-8
20	32	3.3	11049	4	US-10-204-708-22
21	31.8	3.3	828	4	US-09-371-056-9
22	31.8	3.3	1664	4	US-09-397-787-68
23	31.8	3.3	1664	4	US-09-397-787-267
24	31.8	3.3	1892	4	US-09-276-438-12
25	31.8	3.3	4242	4	US-09-276-438-11
26	31.8	3.3	4487	4	US-09-484-970B-56
27	31.8	3.3	44453	4	US-09-146-053-5

28 31.6 3.3 480 4 US-09-621-976-8465 Sequence 8465, Appl
29 31.6 3.3 1185 3 US-08-975-762-15 Sequence 15, Appl
30 31.6 3.3 1185 3 US-08-821-324-15 Sequence 15, Appl
31 31.6 3.3 1185 3 US-09-295-028-15 Sequence 15, Appl
32 31.6 3.3 1185 4 US-09-106-582-15 Sequence 15, Appl
33 31.6 3.3 1185 4 US-09-159-469-15 Sequence 15, Appl
34 31.6 3.3 1185 4 US-09-693-542-15 Sequence 15, Appl
35 31.6 3.3 1185 3 US-08-975-762-39 Sequence 39, Appl
36 31.6 3.3 2129 3 US-09-295-028-39 Sequence 39, Appl
37 31.6 3.3 2129 4 US-09-106-582-39 Sequence 39, Appl
38 31.6 3.3 2129 4 US-09-159-469-39 Sequence 39, Appl
39 31.6 3.3 2129 4 US-09-693-542-39 Sequence 39, Appl
40 31.6 3.3 3998 3 US-09-066-046-5 Sequence 5, Appl
41 31.6 3.3 6376 4 US-09-405-728-1 Sequence 1, Appl
42 31.6 3.3 99500 4 US-09-798-096-10 Sequence 10, Appl
43 31.4 3.3 261 3 US-08-906-769-160 Sequence 160, App
44 31.4 3.3 261 3 US-08-906-616-160 Sequence 160, App
45 31.4 3.3 261 3 US-08-639-075A-160 Sequence 160, App

ALIGNMENTS

RESULT 1
US-09-134-078-60
; Sequence 60, Application US/09134078
; Patent No. 6168844

GENERAL INFORMATION:

APPLICANT: Bylina, Edward J.

TITLE OF INVENTION: GLYCOSIDASE ENZYMES

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Gray Cary Ware & Freidenrich LLP

STREET: 4365 Executive Drive, Suite 1600

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/134,078

FILING DATE: 13-AUG-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/949,026

FILING DATE: 10-OCT-1997

APPLICATION NUMBER: 60/056,916

FILING DATE: 06-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09010/024002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 858/677-1456

TELEFAX: 858/677-1465

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 960 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 1...957

US-09-134-078-60

Query Match 100.0%; Score 960; DB 4; Length 960;
Best Local Similarity 100.0%; Pred. NO. 1.2e-291;

```

Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAGCAAGAAAAGTTGGTTCATCTATCTATCTTAACAATCCCTTTTAGTACAGGCAATA 60
Db 1 ATGAGCAAGAAAAGTTGGTTCATCTATCTATCTTAACAATCCCTTTTAGTACAGGCAATA 60
QY 61 TATTTTGTAGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
Db 61 TATTTTGTAGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA 120
QY 121 CCACCCCAACCAACTTTCCTACTACCAAGTTCTCAAGATTAGATACCTCGATGACGGT 180
Db 121 CCACCCCAACCAACTTTCCTACTACCAAGTTCTCAAGATTAGATACCTCGATGACGGT 180
QY 181 GAGTGCCAGGAGCTTCTTATTGATTAAGATGGTGTGAGGAAACCCGAGATTCTTACATTGAA 240
Db 181 GAGTGCCAGGAGCTTCTTATTGATTAAGATGGTGTGAGGAAACCCGAGATTCTTACATTGAA 240
QY 241 ATAAACCTATGGAACATCTTAACTGCTACTGATTTGCTGAGTACGATCAATTTAAC 300
Db 241 ATAAACCTATGGAACATCTTAACTGCTACTGATTTGCTGAGTACGATCAATTTAAC 300
QY 301 AGCGGCTCTTCACTAGCTCAACCAACTTGACAACTTGTCTTGAGGGATAGAGTAAT 360
Db 301 AGCGGCTCTTCACTAGCTCAACCAACTTGACAACTTGTCTTGAGGGATAGAGTAAT 360
QY 361 TGGTGTCATGGATACCCGGAATATTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
Db 361 TGGTGTCATGGATACCCGGAATATTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
QY 421 ACTGATGGCCCAATACCACTTACCAGTAAAGTTTCAAACTTCAACAGACTTCTTATCA 480
Db 421 ACTGATGGCCCAATACCACTTACCAGTAAAGTTTCAAACTTCAACAGACTTCTTATCA 480
QY 481 ATCTCTATAAATTGAGGCCCAAGAACCGCTGCAATTAATCTGCAATAGATCTCTGG 540
Db 481 ATCTCTATAAATTGAGGCCCAAGAACCGCTGCAATTAATCTGCAATAGATCTCTGG 540
QY 541 TTAACGAGAGAGCTTGGAGAACCAACAGGAATTAACAGGATGACCAAGAGTAATGATA 600
Db 541 TTAACGAGAGAGCTTGGAGAACCAACAGGAATTAACAGGATGACCAAGAGTAATGATA 600
QY 601 TGGATTACTATGACGGATTAACAACCGCTGCTCAAAGTTAAGGAGATTGTAGTCCCA 660
Db 601 TGGATTACTATGACGGATTAACAACCGCTGCTCAAAGTTAAGGAGATTGTAGTCCCA 660
QY 661 ATATAGTTTAAACGACACCACTTAATGCTATGAGTATGAGGCAACCAATTTGGT 720
Db 661 ATATAGTTTAAACGACACCACTTAATGCTATGAGTATGAGGCAACCAATTTGGT 720
QY 721 TGGGAGTATGTTGCAATTTAGATAAAGACCCCAATCAAGAGGGGAACAGTGAATCCA 780
Db 721 TGGGAGTATGTTGCAATTTAGATAAAGACCCCAATCAAGAGGGGAACAGTGAATCCA 780
QY 781 TACGGAGCTTTATAGTCTGACGCAACATTTCAAGCTTACCAATTAACAGCAACTT 840
Db 781 TACGGAGCTTTATAGTCTGACGCAACATTTCAAGCTTACCAATTAACAGCAACTT 840
QY 841 TACTTAGAGGACGTGGAGATTGGAACCTGAGTTTGAACCGCCCAAGCACTACCTCCGCCAC 900
Db 841 TACTTAGAGGACGTGGAGATTGGAACCTGAGTTTGAACCGCCCAAGCACTACCTCCGCCAC 900
QY 901 CTAGAGTGGTGGATCAAAACATTAACACTTCTCTAGATAGACCTCTTATTTCTTAA 960
Db 901 CTAGAGTGGTGGATCAAAACATTAACACTTCTCTAGATAGACCTCTTATTTCTTAA 960

```

RESULT 2

US-08-995-280C-1
; Sequence 1, Application US/08995280C
; Patent No. 6043075
; GENERAL INFORMATION:
; APPLICANT: Bjornvad, Mads Eskelund

```

; APPLICANT: Schulein, Martin
; APPLICANT: NO. 6043075revang, Iben Angelica
; TITLE OF INVENTION: A NO. 6043075el Endoglucanase
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6043075o No. 6043075disk of No. 6043075th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,280C
; FILING DATE: 19-Dec-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol A.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4906.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-995-280C-1

```

Query Match 11.9%; Score 114.2; DB 3; Length 867;

Best Local Similarity 51.3%; Pred. No. 6.6e-26;
Matches 291; Conservative 0; Mismatches 273; Indels 3; Gaps 1;

```

QY 230 TCTACATTGAAATAAAACCTATGGAACATTTCTTAATCTACTGATTTGCTGAGATGACGT 289
Db 137 TCACCTTAGAAGTCAACTTTTGAATATTCGAACCTATGAGGAATACATGATGGCAT 196
QY 290 ACAATTTAACCCAGCGGCTCTTCACCTACGTCCAACTTGCACAACTTGCATTCGAGG 349
Db 197 TTTATAAGAGAAGAGATACCTGTTGAATATTTATCCGACATATAAATACATAGTACTTAAG 256
QY 350 ATAGAGTAATTCGGTGCATGATACCCGAAATATTTCTATGAAACAAAGCCATGGAATG 409
Db 257 ATAAATTCATGGGTACATGGATATCTTGAAGTCTACTATGGGTACAAACCATGGCTG 316
QY 410 CAAACTACGCAACTGATGGCCCAATACCATTTACCCAGTAAAGTTTCAAACCTTAACAGCT 469
Db 317 GCCATGGGAATTCATTTGAGAAATTTAGCTCTTCTTAAAAAGGTATCAGAAATTTCCAGAG 376
QY 470 TCTATCTAACAACTCTCCTATTAACCTTTGAGCCCAAGACCGCTGCCAATTAACCTCGCA 529
Db 377 TTCTCTTCAATCTAAATAACAATATGATGTACGAGAGAAATCTTCTTAATTTTCTTA 436
QY 530 TAGAATCCTGTTAAACAGAGAGAAGCTTTGGAGAACCAACAGGAATTAACAGCGATAGCAAG 589
Db 437 TGGAAACATGGAATAACAAGAACCCCTATCAGAAAACCG ---TTACTTCAGGGGATATAG 493
QY 590 AAGTAATGATGATGATTTACTATGACGGATTAACAACCGCTGCTCCAAAGTTAAGAGA 649
Db 494 AGATGATGGTATGGCTATATGCTAAATAGACTTTCTCTCGAGGGCGAAAGGTAGGAAG 553
QY 650 TTGTAGTCCCAATAATAGTTTAAACGGAACACCACTAAATGCTACATTTGAAGTATGGAAG 709
Db 554 TAAATATACCTATCATCTTAAACCGTAAATCAAAAGACATTTATCTGGGAAGTATATCTTT 613

```


Db 456 TTAATACCAATGAATGATAATCTACTACTAGTAGCAATGTAATGATAGTAAGTAA 397
QY 601 TGGATTACTATGACGGATTAACAACCGCTGGCTCCAAAGTTAAGGAGATTGTAGTCCCA 660
Db 396 ATAATCTTCTTGATAGTTTGGCTCAATTCGATGTAATTAATTAATCAATGTAATCAA 337
QY 661 ATAATAGTTAAGGAAACCAAGTAAATGCTACATTTGAAGTATGGAAGGCAAAATGCT 720
Db 336 GCTACAGATAAAGGAAATAACCAATCCACTTTGCAATGGAGATAATCCTAGTTGTCT 277
QY 721 TGGAGTATGTTGCTTTAGATAAAGACCCCAATCAAGAGGGAACAGTGCACAAATCCA 780
Db 276 TGAAGTAGACTGGAATATATAATTAATGATCCCATTAACAATAGCTGTTAGAGGTCA 217
QY 781 TAGGAGCATTTATAGTGTGACGCCAACATTTCAAGCTTACCAAAATACACAGAA 837
Db 216 GTTATAAAACTAAAGTAATCGAACGATTAAATTCACCCAGGTAATAATGGTGAA 160

RESULT 5

US-09-636-382A-16

; Sequence 16, Application US/09636382A

; Patent No. 6514741

; GENERAL INFORMATION:

; APPLICANT: Presnell, Scott R.

; APPLICANT: Taft, David W.

; TITLE OF INVENTION: TRYPTASE-LIKE POLYPEPTIDE ZTRYPI

; FILE REFERENCE: 99-21

; CURRENT APPLICATION NUMBER: US/09/636,382A

; PRIOR FILING DATE: 2000-08-09

; PRIOR APPLICATION NUMBER: US 60/149,563

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 16

; LENGTH: 936

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Degenerate polynucleotide sequence of human ztryp1

; OTHER INFORMATION: (SEQ ID NO:15)

; NAME/KEY: misc_feature

; LOCATION: (1)...(936)

; OTHER INFORMATION: n = A,T,C or G

US-09-636-382A-16

Query Match 3.8%; Score 36; DB 4; Length 936;
Best Local Similarity 25.8%; Pred. No. 0.26;
Matches 66; Conservative 49; Mismatches 141; Indels 0; Gaps 0;
QY 111 CTCATCTACACCACCCCAACACACTTTCCACTACCAAGTTCTCAAGATTAGATACCC 170
Db 245 CNCAYTGYTNCARMGNTTYAARGAYTNWSNTYNTVWSNGTNGGNGTNCAYC 304
QY 171 TGATGACGGTAGTGCAGGAGCTCTATTGATAGAGATGGTGTGGACCCAGATT 230
Db 305 ARMGNCNGARAAYWSNACNARYTNCNNTYTNACNMGNATGGTNATHCAYARGAYTYW 364
QY 231 CTAATTTGAAATAAACCTATGCAACATTTCTTAATGCTACTACTGATTGCTGAGATGACCTA 290
Db 365 SNAAYTNTATGNSCARGAYATHGNCYNTYNTAARYTNMGAYWSNATHSNATGWSNC 424
QY 291 CAATTTAACCGCGGCTCTTCACTAGCTCCACCACTTGACAACTTGACAACTGCTTGAGGGA 350
Db 425 CNTTYGTNCARCNGTNTGYTTCNCCNAAAYATHAARTTYAARCCNWSNATHGWNATGT 484
QY 351 TAGAAGTAATGSGTG 366
Db 485 GYGGGTATGNGTG 500

RESULT 6

US-09-134-000C-3076/c
; Sequence 3076, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3076
; LENGTH: 1242
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-3076

Query Match 3.6%; Score 34.8; DB 4; Length 1242;

Best Local Similarity 52.0%; Pred. No. 0.71; 72; Indels 0; Gaps 0;

Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 675 AACACCAAGTAAATGCTACATTTGAAGTATGGAAGGCAAAACATTTGGTGGAGTATGTTGC 734

Db 156 AACATCAGACGATTCCTAATATAGAGATCGTTAAACAAAACGTTGTGAATATCTTGTAC 97

QY 735 ATTAGAATAAGACCCGATCAAGAGGGAACAGTGAATCCATCGGAGCATTTAT 794

Db 96 ATTGCCATAAATATCTTCAATCAGCCAGCTTTTCCAAAAATCCATAGCTGCATTTGA 37

QY 795 AAGTGTTCAGGCAACATTTCAAGCTTACC 824

Db 36 AAGTCTATCTCTTAGCTTCAAAATCTCACC 7

RESULT 7

US-09-620-312D-277/c

; Sequence 277, Application US/09620312D

; Patent No. 6569662

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Ren, Feiyan

; APPLICANT: Chen, Rui-hong

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wehrman, Tom

; APPLICANT: Xue, Aidong J.

; APPLICANT: Yang, Yonghong

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhou, Ping

; APPLICANT: Ma, Yunding

; APPLICANT: Wang, Dunrui

; APPLICANT: Wang, Zhiwei

; APPLICANT: John Tillinghast

; APPLICANT: Drmanac, Radoje T.

; TITLE OF INVENTION: No. 6569662el Nucleic Acids and

; TITLE OF INVENTION: Polypeptides

; FILE REFERENCE: 784CIP2B

; CURRENT APPLICATION NUMBER: US/09/620,312D

; CURRENT FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1105

; SOFTWARE: pt_FL_genes Version 1.0

; SEQ ID NO 277

; LENGTH: 3146

; TYPE: DNA

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23)...(3130)
US-09-620-312D-277

Query Match 3.6%; Score 34.8; DB 4; Length 3146;
Best Local Similarity 46.1%; Pred. No. 1.2;
Matches 117; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 287 CGTACAATTAAACGAGCGGCTCTTCACTAGGTCACACACTTGACACACTTCTTGA 346
DB 1503 CTTAAGAAACGAGCGAGAGATGTCTAGAGTCTGGAACCTGCGAAGTTTATTTCA 1444
QY 347 GGGATAGAAAGTAATTTGGGTGCATGGATACCCCGGAAATATTTCTATGGAAACAAAGCCATGGA 406
DB 1443 AGATTGATATCAATTTGGGTGCAGAGAGCTGGGCACTCAGTCACACACCGGCTTGTCT 1384
QY 407 ATGCAAACTAGCAACTGATGGCCCAATACCATTACCCAGTAAAGTTTCAAACTCAAG 466
DB 1383 CTGTAACAGGAGACTTTTCCCTGTTCCAAATGCACCCACAATCAAACTCTGGGTAAATCTTC 1324
QY 467 ACTTCTACTTAACATCTCCTATATAAATTTGAGCCCAAGACGGCTGCCAATTAACTTCG 526
DB 1323 TTGCTATGCTGAATCTCTCTTAAAGTAAAGCAATCCGGAAGGACAGCATGTGAG 1264
QY 527 CAATAGAACTCTGG 540
DB 1263 GCCACACTCCTTG 1250

RESULT 8
US-09-601-198-60/c
; Sequence 60, Application US/09601198
; Patent No. 65311583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 15016
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-60

Query Match 3.6%; Score 34.8; DB 4; Length 15016;
Best Local Similarity 50.6%; Pred. No. 2.6;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY 652 GTAGTCCCAATAAATAGTTAAGGACACACAGTAATGTACATTTGAATGGAAGCA 711
DB 5691 GTTGGGCTGTATCAGTTAATAAACCAGCAAGGAAATTTGATATTCACAGACA 5632
QY 712 AACATTGGTGGAGTAGTTGGCATTTAGATAAAGCCCCCAATCAAGAGGGACAGTG 771
DB 5631 AATCAGCGTTATGATAATTTTATTATATAAAGATATTAATAATCAAAATTTATCTTGA 5572
QY 772 ACAATTCCATACGGAGCATTTATAGTGTTCAGCAACATTTCAA 817
DB 5571 GGAAACATATTTAAAGTTAAAGTATTAATGTTGCACACCCCTTCAA 5526

RESULT 9
US-09-328-352-1845
; Sequence 1845, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Galy L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: CTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1845
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1845

Query Match 3.6%; Score 34.2; DB 4; Length 1161;
Best Local Similarity 54.3%; Pred. No. 1.1;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 717 TGGTTGGAGTATGTTGCATTTAGATATAAGACCCCAATCAAGAGGGGACAGTGACAT 776
DB 78 TGGTGTGGATATTGTTGGCGCTTCATTTGAGCCCTTACAGCTTCATGCCACACTTACAGT 137
QY 777 TCCATACGGAGCATTTTATAAGTGTTCAGGCCAACATTTCAAGCTTACCAAAATTACACAGA 836
DB 138 TCCATTCGAACCGAGTTACGTGATGAGCTTATGGCATTAACTTTACCCGATGACAATGA 197
QY 837 ACTTTAC 843
DB 198 AATTGAC 204

RESULT 10
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: FB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature

```

; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)

```

```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1

```

Query Match 3.5%; Score 33.8; DB 4; Length 1664976;
 Best Local Similarity 53.4%; Pred. No. 59;
 Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 732 TGCATTTAGAATAAGACCCCAATCAAGAGGGAACAGTGCACAACTTCCATACGGAGCAIT 791
Db 274434 TTCACATATGGTATAGGTAAGGAATAATCAGGACAAATAAATCTTTCACATCTCTCAIT 274493
QY 792 TATAAGTGTGACGCCAACATTTCAAGCTTACCAATTAACAGAACTTTTACTTAGAGGA 851
Db 274494 TATAATGTTTAGAGCTTTCATTTGTTGATAAACTATTAAATCAGTTTTTAAATTTATATCT 274553
QY 852 CGTCGAGATTGGA 864
Db 274554 CGTGTAGTTTGA 274566

RESULT 11

US-09-224-024-27/c
; Sequence 27, Application US/09224024
; Patent No. 6056953
; GENERAL INFORMATION:
; APPLICANT: Leslie Hickie
; APPLICANT: Jewel Payne
; TITLE OF INVENTION: Materials and Methods for the Control of
; TITLE OF INVENTION: Calliphoridae Pests
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/224,024
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/856,226
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA79
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-224-024-27

Query Match 3.5%; Score 33.4; DB 3; Length 3543;
Best Local Similarity 46.1%; Pred. No. 3.4;
Matches 148; Conservative 0; Mismatches 171; Indels 2; Gaps 1;

QY 439 TTACCAGTAAGTTTCAAACTTAACAGACTTCTATCAAACTCTCTATAACTTGAG 498
Db 1734 TTGACATGTAATTTTGAATGATCTTTGAATCAATTAATCCCTCTCTGTATGACCAGG 1675
QY 499 CCCAAGAACGGCTGCCAACTTAACCTCGCAATAGAACTCTGTTAACGAGAGAGCTTGG 558
Db 1674 TCCTTGAAACACCTTAGAAGCA--GTCCCAAGTGAATTCGCTTTTACAGCTGGAATTTGG 1617
QY 559 AGACAAAGGAATTAACAGGATGACGACGAAGTAATGATATGATGATTTACTATGACGGA 618
Db 1616 GTAGTTAAATGTGTATAAATTTGTAATTTTAGATCAACACTAGAGTGTGTCCAAAGCAAC 1557

QY 619 TTACAACGGCTGGCTCCAAAGTTAAGAGATTGTAGTCCCAATATATAGTTAACGGAACA 678
Db 1556 GTATACACTTGATTTTATATGTTGCAGGGATACTAAGACTTTTATATAAATGATAAATA 1497
QY 679 COAGTAATATGCTACATTTGAAGTATGGAAGGCAAAACATTTGGTGGAGATATGTTGCAATTT 738
Db 1496 TGACTATAGTTATCATATGTTGAAAAAGGGTAGGGATTGCTTGATTCTCTCTCTGTTTA 1437
QY 739 AGAATAAAGACCCCAATCAA 759
Db 1436 AGAATGTAACCCGAAATA 1416

RESULT 12

PCT-US94-07902-27/c
; Sequence 27, Application PC/TUS9407902
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Street address: 4980 Carroll Canyon Road
; APPLICANT: City: San Diego
; APPLICANT: State/Province: California
; APPLICANT: Country: US
; APPLICANT: Postal code/Zip: 92121
; APPLICANT: Phone number: (619) 453-8030 Fax number: (619) 453-6991
; APPLICANT: Telex number:
; TITLE OF INVENTION: Materials and Methods for the Control of
; TITLE OF INVENTION: Calliphoridae Pests
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07902
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA79
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US94-07902-27

Query Match 3.5%; Score 33.4; DB 5; Length 3543;
Best Local Similarity 46.1%; Pred. No. 3.4;
Matches 148; Conservative 0; Mismatches 171; Indels 2; Gaps 1;

QY 439 TTACCAGTAAGTTTCAAACTTAACAGACTTCTATCAAACTCTCTATAACTTGAG 498
Db 1734 TTGACATGTAATTTTGAATGATCTTTGAATCAATTAATCCCTCTCTGTATGACCAGG 1675
QY 499 CCCAAGAACGGCTGCCAACTTAACCTCGCAATAGAACTCTGTTAACGAGAGAGCTTGG 558
Db 1674 TCCTTGAAACACCTTAGAAGCA--GTCCCAAGTGAATTCGCTTTTACAGCTGGAATTTGG 1617

Qy	559	AGAACACACGAATTAACACGGATGACCAAGAGTAATGATATGGATTTTACTATGACGGA	618
Db	1616	GTAGTTAAATCTGTATATAATTTGTATTTTAGGATCAACACTAGAGTGTGCCAAGCAAAC	1557
Qy	619	TTCAACCGGCTGGCTCCAAAGTTAAAGAGATTGTACTCCCAATAATAGTTTACGGGAACA	678
Db	1556	GTATACACTTGGAGTTTATATGTTTCAGGGATACTAAGACTTTTATAAATGATAAAATA	1497
Qy	679	CCAGTAAATGCTACATTTGAAGTATGGAAGCAACAACTGGTGGGAGTATGTTGCATTT	738
Db	1496	TGACTATATATCATATGTTGGNAAAGGGTAGGGATTGCTTGATTCTCTCTTGGTTTA	1437
Qy	739	AGAATAAAGACCCCAATCAAA	759
Db	1436	AGAAATGGTTAACCCGAAAATA	1416

```

RESULT 13
US-09-385-982-400
; Sequence 400, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON C., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: 11
; FILE REFERENCE: CCDNA-280XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 400
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(612)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-400

```

	Query Match	3.5%	Score 33.2;	DB 3;	Length 612;
	Best Local Similarity	52.1%;	Pred. No. 1.5;		
	Matches	74;	Conservative 0;	Mismatches 68;	Indels 0; Gaps 0;
QY	450	AGTTTCAAACTACAGACTTCTATCTAAACAATCTCTATAAACTTGGAGCCCAAGACGG	509		
DB	30	AGTTCAGACTGTCAAAACATATTCATGTCAAGCTTTTCCAAAGATCAGGCGCTAAAGAAA	89		
QY	510	CCTGCGCAATTAACCTTCGCAATAGAACTCCTGGTTAAACGAGAGAAAGCTTTGGAGAAACAACAGG	569		
DB	90	CRGATCACAAAAGCTTTCATCATCTCCTCATATAACGAGAGAAAGTTTAAAGACAGAGAA	149		
QY	570	AATTAACAGCGATGAGCAAGAA	591		
DB	150	AAGCAACAGTGAATGTGAAGAA	171		

RESULT 14
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: cobert, S. J.Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2

```

; CURRENT APPLICATION NUMBER: US/09/621,976
;
; CURRENT FILING DATE: 2000-07-21
;
; NUMBER OF SEQ ID NOS: 19335
;
; SOFTWARE: Patent.pm
;
; SEQ ID NO 2813
;
; LENGTH: 832
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: 235..399
;
; US-09-621-976-3813

```

	Query Match	3.4%; Score 32.8; DB 4; Length 832;
	Best Local Similarity	9.4%; Pred. No. 2.4;
	Matches 28; Conservative 139; Mismatches 131; Indels 0; Gaps 0	
QY	406 AATGCAAACTACGCAACTGATGCCCAATACCAATTACCCAGTAAAGTTTCAACACTACA	465
Db	MMKMMSTRVCYMCWKCCWGRRCAYWTMARGRVSYAWGKWSMRMSMSMCRTMYVK	249
QY	446 GACTTCATCTCAAAATCTCCTATAAATCTTGACCCCAAGAACGGCGCTGCCAATTAACTTC	525
Db	KGSTYWTMKCTCATCYYWYKWKYKRWMSKTCWSGRGGYMTSYSTRSYMTWAGWMTYM	189
QY	526 CCAATAGAAATCCTCGTTAAACGAGAGAAGCTTGAGAGAACCAACGGAATTAACAGCGATGAG	585
Db	CWMMGRWWSYVYMAWGKKWRYATTWRRAMMMWWAAWTMMYTMWMAWCMSSRGAAYRR	129
QY	586 CAAGAGTAATGATATCGAATTACTATGACGGATTACAACCGCTGGCTCCAAAGTTAAG	645
Db	128 TMMWGVYRWKKSRYRTCAWAYAKTKRSYWCWRKWRKMMMMMAWAKTMMW	69
QY	646 GAGATTGTAGTCCCAATAATAGTTAAACGGAACACCAAGTAAATGCTACATTTGAAATAT	703
Db	68 PACWTRVYRWAWAMRWMMWVMMYVYVYRANKRRMMWRKWSMWWMAWGTWR	11

RESULT 15
US-08-319-704-5
Sequence 5, Application US/08319704
Patent No. 5814617
GENERAL INFORMATION:
APPLICANT: Hoffman, Stephen L.
APPLICANT: Charcoenvit, Yupin
APPLICANT: Hedstrom, Richard C.
APPLICANT: Doclan, Denise L.
TITLE OF INVENTION: Protective 17 kDa Malaria Hepatic and
TITLE OF INVENTION: Erythrocytic Stage Immunogen and Gene
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Naval Medical R & D Command
STREET: Bldg 1, T-12, 8901 Wisconsin Avenue
CITY: Bethesda
STATE: Maryland
COUNTRY: U.S.A
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/319,704
FILING DATE: 07-OCT-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: A. David Spevack
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1422 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-319-704-5

Query Match	3.4%	Score 32.8;	DB 1;	Length 1422;
Best Local Similarity	50.0%	Pred. No. 3.2;	Indels 0;	Gaps 0;
Matches	82;	Conservative 0;	Mismatches 82;	

Qy	675	AACACCGTAATGCTACATTGGAAGTATCGAAGGCAACATTGGTTGGGAGTATGTTGC	734
Db	448	AAAAACAAGTATGGTAAAAAATGGCAAAATATGGCTCCCAAAATGTTATTAAGAGTAAGTTGA	507
Qy	735	ATTTAGATTAAGACCCCAATCAAGAGGGAACAGTGACATTCCTACCGGAGCATTAT	794
Db	508	TTTAAAGTAATATATACTAATTTAAGTATCCATATATACATGTTTATACCTTTTGAGTTAT	567
Qy	795	AAGTGTTCGACCAACATTTCAAGCTTACCAAAATTACACAGAAC	838
Db	568	ACGTAATGTATATACGAACTTGATTTTAATATTTTAAAAAAC	611

Search completed: July 2, 2004, 00:48:11
Job time : 107 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 20:13:10 ; Search time 52 Seconds
(without alignments)
1734.288 Million cell updates/sec

Title: US-09-914-543-46

Perfect score: 1701

Sequence: 1 MSKKFVIVSILTLVQAI.....HLEWITNITLTLDRLPLIS 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	100.0	319	13	US-10-003-759-5
2	1701	100.0	319	13	US-10-121-032-64
3	1701	100.0	319	14	US-10-093-037-64
4	1617	95.1	300	14	US-10-228-063-55
5	464	27.3	274	15	US-10-369-493-3041
6	166.5	9.8	261	13	US-10-003-759-2
7	160.5	9.4	225	14	US-10-294-444-1
8	128	7.5	260	12	US-10-441-625-23
9	128	7.5	260	14	US-10-441-626-23
10	120	7.1	1228	10	US-09-917-384-1
11	120	7.1	1228	9	US-09-917-383-1
12	119.5	7.0	553	9	US-09-888-224-2
13	116	6.8	1043	10	US-09-917-384-6
14	116	6.8	1043	10	US-09-917-383-6
15	114.5	6.7	381	12	US-10-441-625-22

16	114.5	6.7	381	14	US-10-441-626-22	Sequence 22, Appl
17	112.5	6.6	264	12	US-10-441-625-24	Sequence 24, Appl
18	112.5	6.6	264	14	US-10-441-626-24	Sequence 24, Appl
19	111.5	6.6	269	14	US-10-156-761-10105	Sequence 10105, A
20	108.5	6.4	221	10	US-09-917-384-13	Sequence 13, Appl
21	108.5	6.4	221	10	US-09-917-383-13	Sequence 13, Appl
22	108	6.3	371	9	US-09-739-861A-1	Sequence 1, Appl
23	108	6.3	371	9	US-09-795-583-1	Sequence 1, Appl
24	108	6.3	371	12	US-10-441-625-21	Sequence 21, Appl
25	108	6.3	371	14	US-10-441-626-21	Sequence 21, Appl
26	108	6.3	386	9	US-09-739-861A-5	Sequence 5, Appl
27	108	6.3	386	9	US-09-795-583-5	Sequence 5, Appl
28	107.5	6.3	231	10	US-09-917-384-7	Sequence 7, Appl
29	107.5	6.3	231	10	US-09-917-383-7	Sequence 7, Appl
30	107	6.3	228	10	US-09-917-384-14	Sequence 14, Appl
31	107	6.3	228	10	US-09-917-383-14	Sequence 14, Appl
32	103	6.1	708	12	US-09-848-909-29	Sequence 29, Appl
33	103	6.1	842	14	US-10-132-350-50	Sequence 50, Appl
34	103	6.1	842	14	US-10-132-350-52	Sequence 52, Appl
35	103	6.1	1338	12	US-09-848-909-35	Sequence 35, Appl
36	102.5	6.0	533	15	US-10-369-493-6113	Sequence 6113, Ap
37	102.5	6.0	533	15	US-10-369-493-6114	Sequence 6114, Ap
38	100.5	5.9	345	9	US-09-775-195-4	Sequence 4, Appl
39	100.5	5.9	345	14	US-10-155-947-4	Sequence 4, Appl
40	100.5	5.9	345	14	US-10-094-407A-4	Sequence 4, Appl
41	100	5.9	881	9	US-09-850-351A-32	Sequence 32, Appl
42	100	5.9	881	15	US-10-452-002A-8	Sequence 8, Appl
43	99	5.8	259	12	US-10-441-625-5	Sequence 5, Appl
44	99	5.8	259	14	US-10-441-626-5	Sequence 5, Appl
45	99	5.8	4620	12	US-10-282-122A-68921	Sequence 68921, A

ALIGNMENTS

RESULT 1

```
US-10-003-759-5
; Sequence 5, Application US/10003759
; Publication No. US20020102699A1
; GENERAL INFORMATION:
; APPLICANT: Wicher, Kryzysztot B.
; APPLICANT: Holst, Olof Peder
; APPLICANT: Hachem, Maher Yousef Abou
; APPLICANT: Karlsson, Eva Margareta No. US20020102699A1dberg
; APPLICANT: Hreggvidsson, Gudmundur O.
; TITLE OF INVENTION: Thermostable Cellulase
; FILE REFERENCE: P5099PC00
; CURRENT APPLICATION NUMBER: US/10/003,759
; CURRENT FILING DATE: 2001-10-23
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/IS01/00012
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/594,884
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(19)
; OTHER INFORMATION:
; NAME/KEY: DOMAIN
; LOCATION: (28)..(49)
; OTHER INFORMATION: Linker Moiety
; NAME/KEY: DOMAIN
; LOCATION: (50)..(319)
; OTHER INFORMATION: Catalytic Domain
US-10-003-759-5
Query Match 100.0%; Score 1701; DB 13; Length 319;
Best Local Similarity 100.0%; Pred. No. 7.9e-160;
```

Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKFVIVSILTLVQAIYFVKYHTSEDKSTNTSSTPQTTLSTTKVKIRYPDDG 60
DB 1 MSKKFVIVSILTLVQAIYFVKYHTSEDKSTNTSSTPQTTLSTTKVKIRYPDDG 60

QY 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120
DB 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120

QY 121 WHGYPEIFYGNKPNANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180
DB 121 WHGYPEIFYGNKPNANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180

QY 181 LTREAWRTTGINSDEQVEMIIYYDGLQOPAGSKVKEIIVPIIVNGTVPNATFEVWKANIG 240
DB 181 LTREAWRTTGINSDEQVEMIIYYDGLQOPAGSKVKEIIVPIIVNGTVPNATFEVWKANIG 240

QY 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTYELVDEVEIGTEGTPSTTSAH 300
DB 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTYELVDEVEIGTEGTPSTTSAH 300

QY 301 LEWMITNITLPLDRPLIS 319
DB 301 LEWMITNITLPLDRPLIS 319

RESULT 2

US-10-121-032-64
; Sequence 64, Application US/10121032
; Publication No- US20020155550A1

GENERAL INFORMATION:

APPLICANT: Bylina, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Gray Cary Ware & Freidenrich LLP

STREET: 4365 Executive Drive, Suite 1600

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/121,032

FILING DATE: 09-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/134,078

FILING DATE: 13-AUG-1998

APPLICATION NUMBER: 08/949,026

FILING DATE: 10-OCT-1997

APPLICATION NUMBER: 60/056,916

FILING DATE: 06-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09010/024002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 858/677-1456

TELEFAX: 858/677-1465

INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:

LENGTH: 319 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 64:

US-10-121-032-64

Query Match 100.0%; Score 1701; DB 13; Length 319;

Best Local Similarity 100.0%; Pred. No. 7.9e-160;

Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKFVIVSILTLVQAIYFVKYHTSEDKSTNTSSTPQTTLSTTKVKIRYPDDG 60

DB 1 MSKKFVIVSILTLVQAIYFVKYHTSEDKSTNTSSTPQTTLSTTKVKIRYPDDG 60

QY 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120

DB 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120

QY 121 WHGYPEIFYGNKPNANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180

DB 121 WHGYPEIFYGNKPNANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180

QY 181 LTREAWRTTGINSDEQVEMIIYYDGLQOPAGSKVKEIIVPIIVNGTVPNATFEVWKANIG 240

DB 181 LTREAWRTTGINSDEQVEMIIYYDGLQOPAGSKVKEIIVPIIVNGTVPNATFEVWKANIG 240

QY 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTYELVDEVEIGTEGTPSTTSAH 300

DB 241 WEYVAFRIKTPKEGTVTIPYGAFISVAANISSLPNTYELVDEVEIGTEGTPSTTSAH 300

QY 301 LEWMITNITLPLDRPLIS 319

DB 301 LEWMITNITLPLDRPLIS 319

RESULT 3

US-10-093-037-64

; Sequence 64, Application US/10093037

; Publication No- US20030078397A1

; GENERAL INFORMATION:

APPLICANT: Jay M. Short

APPLICANT: Bylina, Edward

APPLICANT: Swanson, Ronald V.

APPLICANT: Mathur, Eric J.

APPLICANT: Lam, David E.

; TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF

; FILE REFERENCE: 09010-024006

; CURRENT APPLICATION NUMBER: US/10/093,037

; CURRENT FILING DATE: 2002-03-06

; PRIOR APPLICATION NUMBER: US 09/910,579

; PRIOR FILING DATE: 2001-07-20

; PRIOR APPLICATION NUMBER: US 09/134,078

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 08/949,026

; PRIOR FILING DATE: 1997-10-10

; PRIOR APPLICATION NUMBER: US 60/056,916

; PRIOR FILING DATE: 1996-12-06

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 64

; LENGTH: 319

; TYPE: PRT

; ORGANISM: Pyrococcus furiosus

US-10-093-037-64

Query Match 100.0%; Score 1701; DB 14; Length 319;

Best Local Similarity 100.0%; Pred. No. 7.9e-160;

Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKFVIVSILTLVQAIYFVKYHTSEDKSTNTSSTPQTTLSTTKVKIRYPDDG 60

DB 1 MSKKFVIVSILTLVQAIYFVKYHTSEDKSTNTSSTPQTTLSTTKVKIRYPDDG 60

QY 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120

DB 61 EWPAGPIDKDGNGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQOLDNIVLDRSN 120

QY 223 VNGTVPNATFVWKNIGWEYVAFRIKTIPIKEGTVITPYGAFISVAANISLPLNYTELYL 282
 DB 184 -----AGATWEVYADWDWNYIAYRRTP-TTSVSELDKAFIDDAVARGYI--RPEWYL 235
 QY 283 EDVEIGTE 290
 DB 236 HAVETGFE 243

RESULT 7
 US-10-294-444-1
 ; Sequence 1, Application US/1029444
 ; Publication No. US20030199072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Crennell, Susan J.
 ; APPLICANT: Karlsson, Eva M. N.
 ; APPLICANT: Hreggvidsson, Gudmundur O.
 ; APPLICANT: Kristjansson, Jakob K.
 ; APPLICANT: Aevansson, Arnthor
 ; TITLE OF INVENTION: CRYSTAL AND STRUCTURE OF A THERMOSTABLE
 ; TITLE OF INVENTION: GLYCOSOL HYDROLASE AND USE THEREOF, AND MODIFIED PROTEINS
 ; FILE REFERENCE: 2739.2009-000
 ; CURRENT APPLICATION NUMBER: US/10/294,444
 ; CURRENT FILING DATE: 2002-11-14
 ; PRIOR APPLICATION NUMBER: IS 6353
 ; PRIOR FILING DATE: 2002-04-19
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 225
 ; TYPE: PRT
 ; ORGANISM: Rhodothermus marinus
 ; US-10-294-444-1

Query Match 9.4%; Score 160.5; DB 14; Length 225;
 Best Local Similarity 31.4%; Pred. No. 2.3e-07;
 Matches 55; Conservative 23; Mismatches 76; Indels 19; Gaps 8;

QY 117 DRSNVHGYPEIFYGNKPNANYATDGPILPSKVSNTDFYLTISYKLEPKNGLPINFA 176
 DB 51 DNGNVAAYPAIFYGCH-WGACTNSG--LPRVQELSD--VRTSWTLTBITGRWNA 104

QY 177 TSWLTREAWRTTGINSDEQVMIWYD-GLQAGSKVKSIIVPIIVNGTVPVNATEVW 235
 DB 105 YDIWSPVTSNGY-SGGAELMIWLNWNGVMPGGSRVATVEL-----AGATWEV 155

QY 236 KANIGWEYVAFRIKTIPIKEGTVITPYGAFISVAANISLPLNYTELYLEDEVEIGTE 290
 DB 156 YADWDWNYIAYRRTP-TTSVSELDKAFIDDAVARGYI--RPEWYLHAVETGFE 207

RESULT 8
 US-10-441-625-23
 ; Sequence 23, Application US/10441625
 ; Publication No. US20030203467A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gualfetti, Peter
 ; APPLICANT: Mitchinson, Colin
 ; APPLICANT: Phillips, Jay Ian
 ; TITLE OF INVENTION: Novel Variant EGIII-Like Cellulase
 ; TITLE OF INVENTION: Compositions
 ; FILE REFERENCE: GC631
 ; CURRENT APPLICATION NUMBER: US/10/441,625
 ; CURRENT FILING DATE: 2003-05-19
 ; NUMBER OF SEQ ID NOS: 64
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 23
 ; LENGTH: 260
 ; TYPE: PRT
 ; ORGANISM: Rhodothermus marinus
 ; US-10-441-625-23

Query Match 7.5%; Score 128; DB 12; Length 260;
 Best Local Similarity 24.1%; Pred. No. 0.00048;
 Matches 61; Conservative 32; Mismatches 110; Indels 50; Gaps 11;

QY 59 DGEWPGAPDKDGDGNPEFYIEI-----NLWNILNATGPAETMYNTLTSGVLHYV 107
 DB 19 DWLFPGDNGKEPEPEPTVELCGFWDARDVAGGRYRVINNVWGAETAQCIIEVGL--- 74

QY 108 QOLDNIVL-----RDRSNVHGYPEIFYGNKPNANYATDGPILPSKVSNTLDFYLTISY 163
 DB 75 -ETGNFTITRADHDNGNVAAYPAIFYGCH-WAPARAIRDCAARAGAVRAHELDVT--- 129

QY 164 KLEPKNGLP-----NFAIESMLTREAWRTTGINSDEQVMIWYD-GLQAGSKVKEI 217
 DB 130 -----PITGRWNAAYDIWFSPTNSNGY-SGGAELMIWLNWNGVMPGGSRVATV 180

QY 218 VVPIIVNGTVPVNATFVWKNIGWEYVAFRIKTIPIKEGTVITPYGAFISVAANISLPLNY 277
 DB 181 EL-----AGATWEVYADWDWNYIAYRRTP-TTSVSELDKAFIDDAVARGYI--R 229

QY 278 TELYLEDEVEIGTE 290
 DB 230 PEWYLHAVETGFE 242

RESULT 9
 US-10-441-626-23
 ; Sequence 23, Application US/10441626
 ; Publication No. US20030186418A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gualfetti, Peter
 ; APPLICANT: Mitchinson, Colin
 ; APPLICANT: Phillips, Jay Ian
 ; TITLE OF INVENTION: No. US20030186418A1el Variant EGIII-Like Cellulase
 ; TITLE OF INVENTION: Compositions
 ; FILE REFERENCE: GC631
 ; CURRENT APPLICATION NUMBER: US/10/441,626
 ; CURRENT FILING DATE: 2003-05-19
 ; NUMBER OF SEQ ID NOS: 64
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 23
 ; LENGTH: 260
 ; TYPE: PRT
 ; ORGANISM: Rhodothermus marinus
 ; US-10-441-626-23

Query Match 7.5%; Score 128; DB 14; Length 260;
 Best Local Similarity 24.1%; Pred. No. 0.00048;
 Matches 61; Conservative 32; Mismatches 110; Indels 50; Gaps 11;

QY 59 DGEWPGAPDKDGDGNPEFYIEI-----NLWNILNATGPAETMYNTLTSGVLHYV 107
 DB 19 DWLFPGDNGKEPEPEPTVELCGFWDARDVAGGRYRVINNVWGAETAQCIIEVGL--- 74

QY 108 QOLDNIVL-----RDRSNVHGYPEIFYGNKPNANYATDGPILPSKVSNTLDFYLTISY 163
 DB 75 -ETGNFTITRADHDNGNVAAYPAIFYGCH-WAPARAIRDCAARAGAVRAHELDVT--- 129

QY 164 KLEPKNGLP-----NFAIESMLTREAWRTTGINSDEQVMIWYD-GLQAGSKVKEI 217
 DB 130 -----PITGRWNAAYDIWFSPTNSNGY-SGGAELMIWLNWNGVMPGGSRVATV 180

QY 218 VVPIIVNGTVPVNATFVWKNIGWEYVAFRIKTIPIKEGTVITPYGAFISVAANISLPLNY 277
 DB 181 EL-----AGATWEVYADWDWNYIAYRRTP-TTSVSELDKAFIDDAVARGYI--R 229

QY 278 TELYLEDEVEIGTE 290
 DB 230 PEWYLHAVETGFE 242

RESULT 10
 US-09-917-384-1

```

; Sequence 1, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: GuxA
US-09-917-384-1

Query Match      7.1%; Score 120; DB 10; Length 1228;
Best Local Similarity 22.7%; Pred. No. 0.028;
Matches 75; Conservative 33; Mismatches 129; Indels 94; Gaps 16;

QY      26 YHTSEDKSTNTASTPQTILSTIKVLKIPYDDGEWPGAPIDKDGGNP----- 75
DB      822 YTVAADVAAGNTSAPSPVTAITTS-----PSPSPPTGTGTTVDTCTPGPNQNGVTSVQG 875
QY      76 -EFVIEINLWN-----ILNATG-FAEMTYNLTSGVLHVQQLDNIVLDRSNWVHG 125
DB      876 DEYRVQTNENSSAQOCLTINTAGTAVTSTANFSGT-----GCAATY 920
QY      126 PEIPYGNKPNNAYADGPIPLSKVKNLTDFTYLTISYKLEPKNGLPINFAISWLTREA 185
DB      921 PSYKKGCHWN---CTTKNKGMPITQISQ-----IGSAVTSWSTTQV 958
QY      186 -----W-----RTTGINSDECEVMWI--YDGLQPGASKVKEIVVPIVNGTPV 228
DB      959 SSGAYDVAYDIWNTSNPTTTG-QPENGTEIMWLNSRGVQPFSGQTATGV-----TVA 1010
QY      229 NATFEVWKA-NIGWEYVAFRIKPIKEGTVTIPIYGAFISVAANISSLPNTYELYLEDVEI 287
DB      1011 GHTWNVWQGSQTSWKIISY-VLTPGATSI NLDLKAIFADAAARGSL--NTSDVLLDVEA 1067
QY      288 GTEF--GTPSTTSAHLEWWTNITLTPLDRP 316
DB      1068 GFIEWQGGQLGNSFSVSVTGTSSTPSP 1098

RESULT 11
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT

```

```

; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS: Unknown
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-09-888-224-2

```

```

Query Match          7.0%; Score 119.5; DB 9; Length 553;
Best Local Similarity 25.1%; Pred. No. 0.0099;
Matches 45; Conservative 22; Mismatches 51; Indels 61; Gaps 9;

Qy      33 STSNTSSPTPTLT-----KVLKIRPPDGE 61
      ||:|||||:||||:||||:||||:
Db      374 TTTTTSPTPTTTTTTSTPTTQTPTTTTTPNNVPEIVNV-LPTSSQ 432
      ||:|||||:||||:||||:||||:

Qy      62 WPGAPIDKGDGN---PEFYIENLNL---NAT-----GFAEMT 96
      ||:|||||:||||:||||:||||:
Db      433 YGTSVEVVCGTQCASSWGAPLWGVKIGNATDNPVWGWDVYKTAQDGTGSK 492
      ||:|||||:||||:||||:||||:

Qy      97 YNLTSGLVHYQQLDNIVLDRSNVHGYPEIFYGNKPNANYATDGP-IPLPKVSML 154
      ||:|||||:||||:||||:||||:
Db      493 MEIRNGVLK-VTNLWNIINHPKYNTM-AYPEVYTGAKPW-GNQPINAPNFIPIKVSQ 548
      ||:|||||:||||:||||:||||:

```

RESULT 13
US-09-917-384-6
; Sequence 6, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HINMER, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment 6
US-09-917-384-6

Query Match	6.8%;	Score 116;	DB 10;	Length 1043;
Best Local Similarity	23.2%;	Pred. No. 0.055;		
Matches	69;	Conservative 31;	Mismatches 105;	Indels 92; Gaps 15;
Qy	34	TNTASTPTQTLLSTTKYKLYRPDDGEWPGAPID-----KDGDNPEFVIE	80	
Db	678	TFTDTGLAAGTAYTYTVAADAGNVSARSPTVDCPTGPNQGVTSVDG-----EYRVQ	733	
Qy	81	INLWN-----INATG-FAETYNLTSGVLHYVOQLDNVLRRSNWVGHYEIPYG	131	
Db	734	TNEWSSAQQCLTINTATGAWTVSTANFSQGT-----GGAPATYPSIYKG	778	
Qy	132	NKPWNANYATDGPILPSPKVNLTDFYLITSYKLEPKNGLPINFATESLWTREA-----	185	
Db	779	CHWGN--CTTKNVGMPIQISQ-----IGSAVTSNITTQVSSGAYD	816	
Qy	186	-----W-----RTTGINSDEQVMIWI-YDGLQPAGSKVKEIWPILVINGCPVNATEV	234	
Db	817	VAYDIWNTSNPTPTTT-QPNGETEIMLNSRGVQPFQSQTATGV-----TVAGHTWNV	868	
Qy	235	WKA-NIGWEYVAFRIKTPKEGTGVTIPYGAFISVAANISLPPNTYLYLEDVEIGTE	290	
Db	869	WGOOOTSXKTSIV-VLTGAGNSINLIDKAFADAAARGSL--NTSDYLLDVEAGPFE	922	

```

RESULT 14
US-09-917-383-6
; Sequence 6, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment 6
; OTHER INFORMATION: GuxA
; US-09-917-383-6

```

```

Query Match      6.8%; Score 116; DB 10; Length 1043;
Best Local Similarity 23.2%; Pred. No. 0.055;
Matches 69; Conservative 31; Mismatches 105; Indels 92; Gaps 15

Qy          34 TGNSTSTPPQTLLSTTKLKIRYDDGEWGPAD-----KDGNGNPEFYIE 80
Db          678 TSFTDTGLAAGFAYTYTVAADAAGNTSAPSTPVDCTPGNQGVTSVQDQ---EYRVQ 733
Qy          81 INLWN-----ILNATG-FAENTYNLTSGVLHYVQQLDNIVLRDSNNWHGYPEIFVG 131
Db          734 TREWSSAQOCLTINTAGMTVTANFSGGT-----GAPATYSIYKG 778
Qy          132 NKFPNANYATDGFIPPLPSKVSNLTD FYLTISYKLEPKNGLPINPAIESMWTREA----- 185
Db          779 CHWGN---CTTKNVGMPIQIQSQ-----IGSAVTSWTTQVSSGAYD 816
Qy          186 ----W----RTTGINGDEBEVMWI--YYDGLQPAGSKVKETWEPIIIVNGTPVNATFEV 234
Db          817 VAYDIWNSPTPTTC-QONGTEIMLNLSRGGVPFGSQATGV-----TVAGHTWNV 868
Qy          235 WKA-NIGWEYVAFRIKTIPEGTWTYPYGAPISVAAINISSLPNYTELIEDVEIGTE 290
Db          869 WOQQOTSMKIISY-VLTFCATSISNLDKAIFADAAAARGSL--NTSDYLDDVEAGPE 922

RESULT 15
US-10-441-625-22
; Sequence 22, Application US/10441625
; Publication No. US20030203467A1
; GENERAL INFORMATION:
; APPLICANT: Gualfetti, Peter
; APPLICANT: Gualtinton, Colin
; APPLICANT: Phillips, Jay Ian
; TITLE OF INVENTION: Novel Variant EGIIII-like Cellulase
; TITLE OF INVENTION: Compositions
; FILE REFERENCE: GC631
; CURRENT APPLICATION NUMBER: US/10/441,625
; CURRENT FILING DATE: 2003-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptomyces lividans CelB
US-10-441-625-22
```

```

RESULT 15
US-10-441-625-22
; Sequence 22, Application US/10441625
; Publication NO. US20030303467A1
; GENERAL INFORMATION:
; APPLICANT: Guaffetti, Peter
; APPLICANT: Matchinson, Colin
; APPLICANT: Phillips, Jay Ian
; TITLE OF INVENTION: Novel Variant EGIII-Like Cellulase
; TITLE OF INVENTION: Compositions
; FILE REFERENCE: GC631
; CURRENT APPLICATION NUMBER: US/10/441,625
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptomyces lividans Ceib
US-10-441-625-22

```




GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2004, 15:30:21 ; Search time 3965 Seconds
(without alignments)
7230.187 Million cell updates/sec

Title: US-09-914-543-45
Perfect score: 960
Sequence: 1 atgagcaagaaaaagttcgt.....atagaacctttatttctctaa 960

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_estc.*
9: gb_estl.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	4.4	1101	29	CNS0182P
2	40.8	4.2	862	29	AG146121 Pan trogl
3	40.6	4.2	1101	29	CNS00GDR
4	40.6	4.2	1201	13	EX394545

5	40.2	4.2	772	12	BI978207
6	40	4.2	560	29	CC799961
7	40	4.2	603	29	CC800184
8	40	4.2	691	29	CC800016
9	40	4.2	885	13	EX425603
10	39.6	4.1	579	14	CB452594
11	39.4	4.1	829	29	CC927838
12	39.4	4.1	966	13	BU519579
13	39	4.1	712	13	EX416727
14	39	4.1	958	29	CG975866
15	38.6	4.0	282	9	AI645051
16	38.6	4.0	416	9	AA656102
17	38.6	4.0	566	13	EX354627
18	38.4	4.0	580	10	AW626092
19	38.4	4.0	588	29	CC800123
20	38.4	4.0	599	29	CL002260
21	38.4	4.0	601	29	CC800157
22	38.4	4.0	641	29	CL002222
23	38.4	4.0	651	29	CL002262
24	38.4	4.0	664	29	CC800120
25	38.4	4.0	669	29	CL002221
26	38.4	4.0	670	29	CL002213
27	38.4	4.0	676	29	CC800124
28	38.4	4.0	698	29	CC800131
29	38.4	4.0	698	29	CL002246
30	38.4	4.0	699	29	CL002205
31	38.4	4.0	699	29	CL002265
32	38.4	4.0	703	29	CL002219
33	38.4	4.0	706	29	CL002227
34	38.4	4.0	729	29	CL002204
35	38.4	4.0	1201	13	EX361080
36	38.2	4.0	513	29	CC800141
37	38.2	4.0	1191	9	AL558073
38	38	4.0	683	29	CC800140
39	38	4.0	742	12	BI434962
40	38	4.0	807	14	CK139957
41	37.8	3.9	704	14	CF806131
42	37.6	3.9	407	14	CD802400
43	37.6	3.9	517	28	BH376219
44	37.6	3.9	675	28	BH265911
45	37.6	3.9	1177	13	EX359598

ALIGNMENTS

RESULT 1
CNS0182P
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

CNS0182P 1101 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence Sp6 end of BAC
BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL108811
AL108811.1 GI:5629115
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CERH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC Project Grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector

COMMENT

```

FEATURES
  source      pBelobAC11.
              1..1101
              Location/Qualifiers
                /organism="Drosophila melanogaster"
                /mol_type="genomic DNA"
                /db_xref="taxon:7227"
                /clone="BACN37D10"
                /clone_lib="DrosBAC"
                /plasmid="pBelobAC11"
                /note="end : Sp6"

ORIGIN
  Query Match      4.4%; Score 42; DB 29; Length 1101;
  Best Local Similarity 17.6%; Pred. No. 2.7;
  Matches 66; Conservative 151; Mismatches 159; Indels 0; Gaps 0;

QY 550 GAAGCTTCGAGAACCAACAGGAATTAACAGCGCATGAGCAAGAGTAATGATGATGATTTAC 609
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 666 GAAADADWGRGRGRGRARARARARARARARARARARARARARARARARARARARARAR 725
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 610 TATGACGAGTACAAACCGCTGGCTCCAAAGTTAAGGAGATTGATGCTCCCAATAATAGTT 669
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 726 DKGCGKATKTAAWAKGRKGTATAWTWDATWATWKAATDWDKRAAAGRRKRDARK 785
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 670 AACGGAACACAGTAATGCTACATTTGAGATGATGGAAGCAACATTTGGTTGGGAGTAT 729
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 786 TARDGRRARTRRRAAWAGRRARAGARRARARARARARARARARARARARARARARARAWT 845
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 730 GTTCGATTTAGATAAGACCCCAATCAAGAGAGGGAACAGTGACAATTCCTATCGGAGCA 789
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 846 TWRDWDWDWDWTDWDDTTAAWDDARARARARARARARARARARARARARARARADDTDK 905
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 790 TTTATAAGTGTTCGACCAACATTTCAAGCTTACCAATACACAGAACTTTACTTAGAG 849
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 906 RWADATTDKDTTKWTTDDDDWDKAKDRWAAKADGAWKWRDRARDWAAATKDDGDKW 965
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 850 GACGTGGAGATGGAACATGAGTTTGGAGCGCAAGCACTACTCTCGCCCACTAGAGTGG 909
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 966 DKWGRGRGKGGKRWKGGTKGKDDDDDKTWTDRDWWWWTTRTKDWWDDGRGRGW 1025
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 910 TGGATCAACAACATNA 925
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1026 TKRWGAWRADAWAR 1041
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
AG146121/c
LOCUS      AG146121      862 bp      DNA      linear      GSS 08-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-007K01.T7, genomic survey
sequence.
ACCESSION AG146121
VERSION    AG146121.1 GI:16675799
KEYWORDS   GSS.
SOURCE     Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
REFERENCE  1
AUTHORS    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      BAC end sequences of Library RPCI-43
JOURNAL    Unpublished
REFERENCE  2
AUTHORS    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbases@sc.riken.go.jp, URL:http://gsp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance

```

of clone tracking errors.

PRIMERS

Sequencing: T7

LIBRARY

Vector : pBacE3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI
 Location/Qualifiers
 1..862
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="RP43-007K01.T7"
 /sex="male"
 /cell_type="lymphocytes"
 /clone_lib="RPCI-43 Chimpanzee Male BAC Library"

ORIGIN

```

  Query Match      4.2%; Score 40.8; DB 29; Length 862;
  Best Local Similarity 48.7%; Pred. No. 5.1;
  Matches 111; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 570 AATTAAACGCGATGAGCAAGTAATGATATGGATTACTATGACGGATTACACCGGC 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 803 AAATAACATCAACAACAAGATGATAGAACATAGACACGGTTAACTGAGCGGAGGCA 744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 630 TGGCTCCAAAGTTAAGGAGATTGTAGTCCCAATATAGTTAACGGAACACAGTAATGC 689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 743 TGAAGAAGAGAGAGAGAGAAACAAGAACAAAGAAAGATTAAGAAGACAGATTAAGAAG 684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 690 TACATTGTAAGTATGGAAGCAACATTTGGTTGGAGTATGTTGATTAGATAAAGAC 749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 683 GAACACAGCAACATGGAATAATAAGAGTAGTAAAGACATGATGATGGAAGGACAA 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 750 CCCATCAAGAGGAAAGTGACATTCATCCATACGGAGCAATTATAAG 797
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 AAGAGGAGTAGAAGGAGCAGTAAACATAAACAAGCAATCAATAG 576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 3

CNS00GDR

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CNS00GDR 1101 bp DNA linear GSS 03-JUN-1999
 Drosophila melanogaster genome survey sequence T7 end of BAC:
 BACR33M19 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org>. The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Oseogawa and
 Aaron Mammoss in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 F1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be

found at <http://bacpac.med.buffalo.edu/drosophila bac.htm>.

```

FEATURES
source
round at http://pacapack.berkeley.edu/mosophila2_bac.new.
Location/Qualifiers
1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR33M19"
/clone.lib="RPI-98"
/note="end : 77"

ORIGIN
Query Match 4.2%, Score 40.6; DB 29; Length 1101;
Best Local Similarity 18.8%; Pred. No. 6.5;
Matches 65; Conservative 125; Mismatches 156; Indels 0; Gaps 0;

QY 232 TACATTGAATAAACCTATGGAACATCTCTTAATGCTACTGATGCTTGTGAGATGACGTAC 291
DB 738 TWGTTTMTWKATGTMKCKKAWKWTATAWWARMTAMAWDATRTTAAATAATTTDDWTAK 797
QY 292 AATTAAACCAAGCGGCGTCTTCACTAGTCCAAACAACCTTGCAACAATTTGTTGAGGAT 351
DB 798 TTTCCTRMVMMCHGMAAATKKKGAMTRACWKGARWRAKAAKAAWKAAMKATKKKWKAA 857
QY 352 AGAAGTAAATGGGTGCATGNTATCCCGAATATCTTATGGAACAAGCCATGGAATGCA 411
DB 858 DKMKATRMKSKCTGKAWKMTAMRTTKRATDMTTKAAKGAAMKCATTAARAANAM 917
QY 412 AACTACCAACTGATGGCCCAATACCAATTACCCAGTAAAGTTTCAAACCTTAACAGACTTC 471
DB 918 WKAMATGKATGWKADDDANKRWKKKBKCKKAKKADDDKKBHMGCRCTAAMTKAWAKR 977
QY 472 TATCTACAATCTCTTAATACTTGAGCCCAAGAAGCGCTGCCAATTAACCTTGCATA 531
DB 978 AAADMTKKDANKAAAKGMMNMMKMKKKKAKAWKATGCKAMMAAKMDKKKWCCKKA 1037
QY 532 GAATCTCTGGTTAAACAGAGAGCTTGAGAAACAACAGGAATAACA 577
DB 1038 DCMTMTKADMAWKKAADAYDDAKDKRMENRAAMKAAAKAMM 1083

RESULT 4
BX394545
LOCUS
DEFINITION
BX394545 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DC017YB23 5-PRIME, mRNA sequence.
ACCESSION
BX394545
VERSION
BX394545.1 GI:30616404
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DC017CA12Q91.
Location/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC017HB23"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone.lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-cligo(dT)
FEATURES
source

```

```

/clone_lib="Old Blush petal SMART library"
/notes="Organ: Petal; Vector: pTriplex2; Site_1: Sfil;
Site_2: Sfil"

ORIGIN

Query Match      4.2%; Score 40.2; DB 12; Length 772;
Best Local Similarity 49.3%; Pred. No. 7.2;
Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 587 AAGAGTAATGATGATGATTTACTATGACGATTACAAACGGCTGGCTCCAAAGTTAAGG 646
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 AAGGAGATTGGTGAATTAACATATGATGATGCTGCGACAAATGGTATTTGTTGATA 399
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 647 AGATTGTAGTCCCAATATAGTTTACGGAACACACAGTAATGCTPACATTTGAAAGTATGGA 706
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 GCACGTGTTGATCAATTCATGAGAAAGACACACCTGTGATCTTCCTTTGGACGAAATC 459
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 707 AGGCAACATTTGGTGGAGTATGTCATTTAGTAATAAGACCCCAATCAAGAGGGAA 766
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 AGATTATAAAGGCTTGGAGGAGGTTCTGTTGGCATGAGAAATCGAGGAAAGAGGAGAG 519
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 767 CAGTCACAATTCATACGAGCATTTTAAAGT 799
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 CATTGATACCTCCTCTGTGGATACACAAATG 552
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
CC799961
LOCUS
DEFINITION
0250069-08A1-A09 UniformMu MutAIL Library Zea mays genomic clone
ACCESSION
CC799961
VERSION
CC799961.1 GI:32622384
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 560)
Latshaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.
Sequence tagged transposon insertions from the UniformMu maize
population
Unpublished (2003)
CONTACT: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmc@ufl.edu
Sequence flanking probable Mu insertion site in UniformMu maize
0250069-08, Primer set: A
Class: transposon insertion site.
Location/Qualifiers
1..560
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="0250069-08A1-A09"
/clone_lib="UniformMu MutAIL Library"
/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon
insertions in Mu inactive lines were extracted from the
UniformMu maize population by the thermo asymmetric
interlaced PCR (TAIL) protocol using primers specific for
the Mu terminal inverted repeat and a set of 16 arbitrary
primers. Amplicons were size enriched using Sepharose 400
spin columns and cloned into the TOPO PCR4 vector."

FEATURES
source
Query Match      4.2%; Score 40; DB 29; Length 560;
Best Local Similarity 53.1%; Pred. No. 7.1;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 786 AGCATTTTAAGTGTGCGACCAACATTTCAAGCTTACCAATTTACACAGAACTTTACTT 845
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 AACATTTTAAACGACACACTGAAAATTTACAACTTTTATTAATGGCTCACAACATTGCTT 437
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 846 AGAGACGTGGAGATGGAACTGAGTTTGGACGCCAAGCACTACTCGGCCACCTAGA 905
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 ATAAGAGGTACAGAGTATAAAACAGAAATGCTTGGAGAGTACTCCCTCCGTTCTTTTAA 497
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 906 GTGGTGGATCAAAAACATAACACTAACTCCTCTAGATAGA 945
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 786 AGCATTTTAAGTGTGCGACCAACATTTCAAGCTTACCAATTTACACAGAACTTTACTT 845
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 AACATTTTAAACGACACACTGAAAATTTACAACTTTTATTAATGGCTCACAACATTGCTT 435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 846 AGAGACGTGGAGATGGAACTGAGTTTGGACGCCAAGCACTACTCGGCCACCTAGA 905
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 ATACGAGGTACAGAGTATCAAAACAGAAATGCTTGGAGAGTACTCCCTCCGTTCTTGCGA 495
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 906 GTGGTGGATCAAAAACATAACACTAACTCCTCTAGATAGA 945
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 TTTGTCGTGATAGCGTAAATTCACACTATCCAGCGACA 535
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
CC800184
LOCUS
DEFINITION
0250069-08D1-C08 UniformMu MutAIL Library Zea mays genomic clone
ACCESSION
CC800184
VERSION
CC800184.1 GI:32622727
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 603)
Latshaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.
Sequence tagged transposon insertions from the UniformMu maize
population
Unpublished (2003)
CONTACT: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmc@ufl.edu
Sequence flanking probable Mu insertion site in UniformMu maize
0250069-08, Primer set: D
Class: transposon insertion site.
Location/Qualifiers
1..603
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="0250069-08D1-G08"
/clone_lib="UniformMu MutAIL Library"
/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon
insertions in Mu inactive lines were extracted from the
UniformMu maize population by the thermo asymmetric
interlaced PCR (TAIL) protocol using primers specific for
the Mu terminal inverted repeat and a set of 16 arbitrary
primers. Amplicons were size enriched using Sepharose 400
spin columns and cloned into the TOPO PCR4 vector."

FEATURES
source
Query Match      4.2%; Score 40; DB 29; Length 603;
Best Local Similarity 53.1%; Pred. No. 7.3;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 786 AGCATTTTAAGTGTGCGACCAACATTTCAAGCTTACCAATTTACACAGAACTTTACTT 845
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 AACATTTTAAACGACACACTGAAAATTTACAACTTTTATTAATGGCTCACAACATTGCTT 437
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 846 AGAGACGTGGAGATGGAACTGAGTTTGGACGCCAAGCACTACTCGGCCACCTAGA 905
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 ATAAGAGGTACAGAGTATAAAACAGAAATGCTTGGAGAGTACTCCCTCCGTTCTTTTAA 497
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 906 GTGGTGGATCAAAAACATAACACTAACTCCTCTAGATAGA 945
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

RESULT 10
LOCUS      CB452594
DEFINITION 707465 MARC 6BOV Bos taurus cDNA 5', mRNA linear EST 26-MAR-2003
ACCESSION  CB452594
VERSION     CB452594.1 GI:29258976
KEYWORDS   EST.
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE  1 (bases 1 to 579)
AUTHORS   Smith, T.P.L., Roberts, A.J., Echtenkamp, S.E., Chitko-McKown, C.G.,
            Wray, J.E. and Keele, J.W.
TITLE     A second set of bovine ESTs from pooled-tissue normalized libraries
JOURNAL   Unpublished (2003)
COMMENT   USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: F0Y8064 row: C column: 13
            Seq primer: GTAATACGACTCACTATAGG.
FEATURES   source
            Location/Qualifiers
                1..579
                /organism="Bos taurus"
                /mol_type="mRNA"
                /db_xref="taxon:9913"
                /tissue_type="pooled"
                /lab_host="DH10B"
                /clone_lib="MARC 6BOV"
                /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
                Library made with RNA pooled from multiple tissues
                including liver, lung, hypothalamus, pituitary, and
                placenta/endometrium."
ORIGIN
Query Match      4.1%; Score 39.6; DB 14; Length 579;
Best Local Similarity 54.9%; Pred. No. 9.3;
Matches 78; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 694 TTGAGATGGAAGGCAACATTTGGTGGAGATGTTGCATTTAGAATAAGACCCCA 753
Db 297 TTTTAATTCCTCAATCTTACTTCTGTGGAATATAAGCGTTTGAACTTTAATACAAG 356

QY 754 ATCAAGAGGGAACGTGACATTCCTACATGACGACATTTATAGTGTTCGAGCAACATT 813
Db 357 TTAAGAAATTTAACAATTAATTTTCAATGACCATTTTAAAGTCCTTAAGACAAATTT 416

QY 814 TCAAGCTTACCAATTTACACAG 835
Db 417 CTTAACTTTTCTATTGCACAG 438

RESULT 11
LOCUS      CC927838/c
DEFINITION CC927838 Zea mays subsp. mays genomic clone
ACCESSION  CC927838
VERSION     CC927838.1 GI:33578936
KEYWORDS   GSS.
SOURCE     Zea mays subsp. mays (maize)
ORGANISM   Zea mays subsp. mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 829)

```

```

AUTHORS   Bharti, A.K., Young, S., Kavchok, S., Keizer, G., Bronzino, A.C.,
            Rouzard, K., Fuks, G., Yu, Y., Wing, R. and Messing, J.
TITLE     Sequencing of the maize genome at PGIR (2003b)
JOURNAL   Unpublished (2003)
COMMENT   Contact: Bharti, A.K.
            Dr. Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 102.
FEATURES   source
            Location/Qualifiers
                1..829
                /organism="Zea mays subsp. mays"
                /mol_type="genomic DNA"
                /cultivar="B73"
                /sub_species="mays"
                /db_xref="taxon:4578"
                /clone="ZMMBBc0530M15"
                /lab_host="E. coli DH10B"
                /clone_lib="ZMMBBc"
                /note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"
ORIGIN
Query Match      4.1%; Score 39.4; DB 29; Length 829;
Best Local Similarity 53.6%; Pred. No. 12;
Matches 82; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 661 ATAATAGTTAACGGAACCCAGTAAATGCTCAATTTGAAGTATGGAAGGCAAACTTGGT 720
Db 584 ATTTTAGACTATAAAATTTATATGTCATATAAAATGTTTGGATGTCAAATATAAGT 525

QY 721 TGGGAGTATGTCATTTAGATAAGACCCCAATCAAGAGGGAACAGTCACCAATTCGA 780
Db 524 ACTAATATAGTCTAATTTATATAAAATATTCATCGATTAAGACTAAAAGACATTTT 465

QY 781 TACGAGCATTTATAGTGTTCAGCCCAACT 813
Db 464 TAAGATAATTAATTTGTGGAGGCAAAATATT 432

RESULT 12
LOCUS      BUS19579/c
DEFINITION BUS19579 AGENCOURT 10166786 NIH MGC 134 Mus musculus cDNA clone
ACCESSION  BUS19579
VERSION     BUS19579.1 GI:22827105
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 966)
            NIH-MGC http://mgi.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dr. David Rowe
            CDNA Library Preparation: Invitrogen Corp
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILLNL at:
            http://image.llnl.gov
            Plate: LLAM14097 row: i column: 08
            High quality sequence start: 62
            High quality sequence stop: 563.

```

FEATURES source Location/Qualifiers
 1..966
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6517423"
 /tissue_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_134"
 /note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
 Cloned unidirectionally. Primer: Oligo dT. Average insert
 size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
 this is a NIH_MGC Library."

ORIGIN

Query Match 4.1%; Score 39.4; DB 13; Length 966;
 Best Local Similarity 52.8%; Pred. No. 13;
 Matches 85; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
 QY 218 GGAACCCAGATCTTACATTGAATAAATCTATGGAACATCTTAATGCTACTGGATTG 277
 Db 165 GCGAAAAGACTTGAATAATTGGAAGGTAAACATTGAAAATTAATAAGGTTTATTATTT 106
 QY 278 CTGAGATGAGCTACAAATTTAACGACGGCGTCTTCTACTAGTCCAAACAATTTGACAACA 337
 Db 105 AAAAAAAGGAAAAAAGGCGGCCATCCTGAATTCGGACCGTACCAGCC 46
 QY 338 TTGCTTTGAGGATAGAAATTAATGGGTGATGATACCC 378
 Db 45 TCTTTGGGGGTTGGCGGTGGATGGTGGGAAGAAGGCC 5

RESULT 13
 BX416727/c
 LOCUS BX416727 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
 DEFINITION CSODA011Y114 5-PRIME, mRNA sequence.
 ACCESSION BX416727
 VERSION BX416727.1 GI:30765629
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CSODA011BE07QPI.

FEATURES source Location/Qualifiers
 1..712
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSODA011Y114"
 /tissue_type="NEUROBLASTOMA"
 /clone_lib="Homo sapiens NEUROBLASTOMA"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN

Query Match 4.1%; Score 39; DB 13; Length 712;
 Best Local Similarity 10.0%; Pred. No. 15;

Matches 31; Conservative 147; Mismatches 133; Indels 0; Gaps 0;
 QY 290 ACAATTTAACACGCGCGTCTTCTACTAGTCCAAACAATTTGCTTTGAGGG 349
 Db 693 MMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMM 634
 QY 350 ATAGAAGTAATTTGGGTGATATACCCGGAATATTTCTATGGAACAAGCCATGGAATG 409
 Db 633 MMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMM 574
 QY 410 CAAACTAGCAACTGATGCCCAATACCATACCCAGTAAAGTTTCAAAACCTAACAGACT 469
 Db 573 MMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMM 514
 QY 470 TCTATCTAAATCTCTTATAAATTTGAGCCCAAGACGGCTGCCAATTAATCTCGCAA 529
 Db 513 MMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMM 454
 QY 530 TAGAATCCTGTTTACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAAG 589
 Db 453 AAAAAAAMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMMMTMM 394
 QY 590 AAGTAATGATA 600
 Db 393 AAAAAAATAA 383
 CG975866 958 bp DNA linear GSS 15-DEC-2003
 MBEAX64TR mth2 Medicago truncatula genomic clone 16L8, genomic
 survey sequence.
 ACCESSION CG975866
 VERSION CG975866.1 GI:39901645
 KEYWORDS GSS.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 REFERENCE 1 (bases 1 to 958)
 AUTHORS Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.
 TITLE Sequencing of BAC ends from Medicago truncatula
 JOURNAL Unpublished (2003)
 COMMENT Other GSSs: MBEAX64TFB
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 Seq primer: CAGGAACAGCTATGACC
 Class: BAC ends.
 FEATURES Location/Qualifiers
 source 1..958
 /organism="Medicago truncatula"
 /mol_type="genomic DNA"
 /cultivar="genotype A1.7"
 /db_xref="taxon:3880"
 /clone="16L8"
 /clone_lib="mth2"
 /note="Vector: pBeloBAC11; Site 1: HindIII; Site 2:
 HindIII; Cook, D.R. and Kim, D.J, unpublished"

ORIGIN

Query Match 4.1%; Score 39; DB 29; Length 958;
 Best Local Similarity 59.5%; Pred. No. 17;
 Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
 QY 4 AGCAAGAAAAAGTTCGTCATCTATCTTAAACAATCTTTAGTACAGGCAATATAT 63
 Db 409 ATCAATAAAAAATTATGTTATTTATTTATTTGAAATTTATTTATCAAAATACATTAATT 468

Search completed: July 2, 2004, 00:46:15
Job time : 3969 secs

```

QY      64 TTTGTAGAAAGTATCATACCTCTGAGGACAAAGTCAACTTCAAAATACCTCA 114
      |||||
Db      469 TTTGTGTATATCTAATGCGTATATATAGAGTATTATATATATATACATCA 519

RESULT 15
AI645051      282 bp      mRNA      linear      EST 29-APR-1999
LOCUS      vs46e05.v1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone
DEFINITION      IMAGE:1149344 5', mRNA sequence.
ACCESSION      AI645051
VERSION      AI645051.1 GI:4723526
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      1 (bases 1 to 282)
      Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
      Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
      Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
      Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
      Waterston,R. and Wilson,R.
      The WashU-NCI Mouse EST Project 1999
      Unpublished (1999)
      Contact: Marra M/WashU-NCI Mouse EST Project 1999
      Washington University School of Medicine
      4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
      Tel: 314 286 1800
      Fax: 314 286 1810
      Email: mouseest@watson.wustl.edu
      This clone is available royalty-free through LNL ; contact the
      IMAGE Consortium (info@image.llnl.gov) for further information.
      MGI:622552
      This read is a RESEQUENCE of a previously sequenced mouse clone
      This read has been verified (found to hit its original self in the
      correct orientation)
      Seq primer: -40RP from Gibco
      High quality sequence stop: 273.
      Location/Qualifiers
        1..282
          /organism="Mus musculus"
          /mol_type="mRNA"
          /db_xref="taxon:10090"
          /clone="IMAGE:1149344"
          /tissue_type="Tcell"
          /dev_stage="M30 CD4+ cells"
          /lab_host="SOLR (kanamycin resistant)"
          /clone_lib="Stratagene mouse Tcell 937311"
          /note="Organ: blood; Vector: pBluescript SK-; Site 1:
          EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
          Oligo dT. M30 CD4+ cells. Average insert size: 1.0 kb;
          Uni-ZAP XR vector; -5' adaptor sequence: 5' GAATTCGGCAGG
          3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

ORIGIN
      Query Match      4.0%; Score 38.6; DB 9; Length 282;
      Best Local Similarity 56.8%; Pred. No. 13;
      Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY      25 GTATCTATCTTAACAACTCTTTTAGTACAGCAATATATTTGTAGAAAAGTATCATACC 84
      |||||
Db      57 GAATGTATAAAAACCTATCTCTTTGTGTAAAGCAACCACTTTGTGAAGCAGACAGCAATACA 116

QY      85 TCTGAGGACAGTCAACTCAATACCTCATCTACACCCCAACAAACACACTTCCACT 144
      |||||
Db      117 CACAATGACATATCATCTATCCCAATGACAGAGTCCTCCTGTTAGACACACTGAAGT 176

QY      145 ACCAA 149
      |||||
Db      177 AGCAA 181
  
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 19:53:43 ; Search time 61 Seconds
(without alignments)
1477.584 Million cell updates/sec

Title: US-09-914-543-46

Perfect score: 1701
Sequence: 1 MSXXKFVIVSILRILLVQAI.....HLEWITNITLTPLDRLPLIS 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Geneseqp29Jan04.*
- 2: Geneseqp1980s.*
- 3: Geneseqp1990s.*
- 4: Geneseqp2000s.*
- 5: Geneseqp2001s.*
- 6: Geneseqp2002s.*
- 7: Geneseqp2003as.*
- 8: Geneseqp2003bs.*
- 9: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	100.0	319	2	AAW35007 Pyrococcus
2	1701	100.0	319	2	AAW29729 Pyrococcus
3	1701	100.0	319	2	AAW49875 Pyrococcus
4	1701	100.0	319	2	AAW23155 Pyrococcus
5	1701	100.0	319	5	AAW50997 Pyrococcus
6	1701	100.0	319	5	AAW50997 Pyrococcus
7	1701	100.0	319	7	AAW50997 Pyrococcus
8	1617	95.1	300	6	ADCP96635 Pyrococcus
9	486.5	28.6	288	2	AAW23156 Dictyoclo
10	480	28.2	297	2	AAW57777 Recombina
11	480	28.2	297	2	AAW29730 Endo-beta
12	439.5	25.8	332	2	AAW17865 Sulfolobu
13	430	25.3	332	2	AAW17866 Sulfolobu
14	328	19.3	841	2	AAW34985 Archaeabac
15	166.5	9.8	243	5	AAW50977 Truncated
16	166.5	9.8	244	5	AAW50975 Truncated
17	166.5	9.8	261	5	AAW50975 Rhodother
18	166.5	9.8	261	5	AAW50975 Rhodother
19	165.5	9.7	240	5	AAW50980 Truncated
20	165.5	9.7	241	5	AAW50979 Truncated
21	165.5	9.7	242	5	AAW50978 Truncated
22	165.5	9.7	628	2	AAW34999 Archaeabac
23	164.5	9.7	239	5	AAW50981 Truncated
24	163.5	9.6	233	5	AAW50987 Truncated
25	163.5	9.6	234	5	AAW50986 Truncated

26	163.5	9.6	235	5	AAW50985 Truncated
27	163.5	9.6	236	5	AAW50984 Truncated
28	163.5	9.6	237	5	AAW50983 Truncated
29	163.5	9.6	238	5	AAW50982 Truncated
30	160.5	9.4	224	5	AAW50996 Truncated
31	160.5	9.4	225	5	AAW50995 Truncated
32	160.5	9.4	226	5	AAW50994 Truncated
33	160.5	9.4	227	5	AAW50993 Truncated
34	160.5	9.4	228	5	AAW50992 Truncated
35	160.5	9.4	229	5	AAW50991 Truncated
36	160.5	9.4	230	5	AAW50990 Truncated
37	160.5	9.4	231	5	AAW50989 Truncated
38	160.5	9.4	232	5	AAW50988 Truncated
39	143.5	8.4	261	2	AAW89462 Bacillus
40	132.5	7.8	429	3	AAW84346 Amino aci
41	128	7.5	260	3	AAW06369 Rhodother
42	128	7.5	260	3	AAW84347 Amino aci
43	128	7.5	260	3	AAW14882 Emericell
44	128	7.5	260	3	AAU77590 R. marinu
45	128	7.5	260	5	AAU77434 Rhodother

ALIGNMENTS

RESULT 1

AAW35007
ID AAW35007 standard; protein; 319 AA.

XX AAW35007;

AC AAW35007;

DT 17-OCT-2003 (revised)

DT 21-MAY-1998 (first entry)

XX

DE Pyrococcus furiosus endoglucanase.

XX

KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;
KW beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase.

OS Pyrococcus furiosus; (Clone 7EG1).

XX WO9744361-A1.

XX 27-NOV-1997.

XX 22-MAY-1997; 97WO-US008793.

XX 22-MAY-1996; 96US-00651572.

XX (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX Lam DE, Mathur BJ;

XX WPI; 1998-018435/02.

XX N-PSDB; AAT94215.

DR Endo:glucanase(s), preferably form archaeal bacterium, ABPII 1a - useful
PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic
PT bonds in cellulose.

XX Claim 1; Fig 1W; 164pp; English.

CC This protein comprises an endoglucanase of Pyrococcus furiosus (Clone
7EG1), that is capable of degrading carboxymethylcellulose and of
hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has homology
to an endoglucanase of archaeobacterium ABPIIa (see AAW34983). It can be
produced from native cells or from recombinant host cells, especially
prokaryotic host cells transformed with a plasmid or virus-derived vector
including the endoglucanase DNA (see AAT94215). 24 Endoglucanases (see
CC AAW34986-W35008) are claimed. They can be used to degrade cellulose for
the conversion of plant biomass into fuels and chemicals, for use in
detergents, textiles, animal feed, waste treatment, and in the fruit

CC juice and brewing industries for the clarification and extraction of
CC juices. (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 319 AA;
Query Match 100.0%; Score 1701; DB 2; Length 319;
Best Local Similarity 100.0%; Pred. No. 3.1e-150;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MSKKKFEIVSILTLVQAIYFVEKYHTSEDKSTNSSTPQTTLSTTKVKIRYPDDG 60
Db 1 MSKKKFEIVSILTLVQAIYFVEKYHTSEDKSTNSSTPQTTLSTTKVKIRYPDDG 60
Qy 61 EWPAGAPIDKGDGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120
Db 61 EWPAGAPIDKGDGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120
Qy 121 WHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180
Db 121 WHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180
Qy 181 LTREAWRTTGINSDEQEWMIYIDGLQAGSKVKEIWPVPIIIVNGTVPVATFEVWKANIG 240
Db 181 LTREAWRTTGINSDEQEWMIYIDGLQAGSKVKEIWPVPIIIVNGTVPVATFEVWKANIG 240
Qy 241 WEYVAFRIKTPKEGTVTIPYGAFTSVAANISSLPNTYLEDVEIGTFGTPSTTSAH 300
Db 241 WEYVAFRIKTPKEGTVTIPYGAFTSVAANISSLPNTYLEDVEIGTFGTPSTTSAH 300
Qy 301 LEWMITNITLPLDRPLIS 319
Db 301 LEWMITNITLPLDRPLIS 319
RESULT 2
AAW29729
ID AAW29729 standard; protein; 319 AA.
XX
AC AAW29729;
XX
DT 18-NOV-1998 (first entry)
DE Pyrococcus furiosus endo-beta-1, 4-glucanase.
KW Pyrococcus furiosus endo-beta-1, 4-glucanase; cellulolytic activity;
KW textile industry; cellulosic fibre; in industrial cleaning process;
KW sugar; instant coffee; oil industry; hydrocolloid cellulose derivative;
KW drilling.
XX
OS Pyrococcus furiosus.
XX
PN WO9833895-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-DK000039.
XX
PR 31-JAN-1997; 97DK-00000114.
PR 11-JUL-1997; 97DK-00000853.
XX
XX (NOVO) NOVO-NORDISK AS.
XX
XX Andersen L, Bjornvad MB, Schuelein M;
XX
XX WPI; 1998-437450/37.
XX
XX N-PSDB; AAV47540.
XX
XX Isolated endo-beta-1,4-glucanase - used for e.g. treating cellulosic
XX fibres or polymers, feed production or in oil industry for enhancing oil
XX recovery.
XX
XX Disclosure; Page 45-46; 56pp; English.

CC The present sequence represents the Pyrococcus furiosus endo-beta-1, 4-
CC glucanase (EG) protein sequence. The invention provides for an enzyme
CC composition having EG activity which has optimum activity at a
CC temperature of at least 90 deg. C. The EG enzyme exhibit cellulolytic
CC activity at extremely high temperatures in a very broad pH range.
CC Therefore it is claimed to be useful for, e.g. in the textile industry
CC for improving the properties of cellulosic fibres or fabric; for
CC providing a stone-washed look of denim; in industrial cleaning processes;
CC in the conversion of biomass to sugars; in the production of instant
CC coffee or similar extraction processes or in the oil industry for
CC degradation of aqueous solutions of hydrocolloid cellulose derivatives
CC used in drilling
XX
SQ Sequence 319 AA;
Query Match 100.0%; Score 1701; DB 2; Length 319;
Best Local Similarity 100.0%; Pred. No. 3.1e-150;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MSKKKFEIVSILTLVQAIYFVEKYHTSEDKSTNSSTPQTTLSTTKVKIRYPDDG 60
Db 1 MSKKKFEIVSILTLVQAIYFVEKYHTSEDKSTNSSTPQTTLSTTKVKIRYPDDG 60
Qy 61 EWPAGAPIDKGDGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120
Db 61 EWPAGAPIDKGDGNPEFYIEINLNLNATGFAEMTNLTSGVLHYVQQLDNIVLRDRSN 120
Qy 121 WHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180
Db 121 WHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180
Qy 181 LTREAWRTTGINSDEQEWMIYIDGLQAGSKVKEIWPVPIIIVNGTVPVATFEVWKANIG 240
Db 181 LTREAWRTTGINSDEQEWMIYIDGLQAGSKVKEIWPVPIIIVNGTVPVATFEVWKANIG 240
Qy 241 WEYVAFRIKTPKEGTVTIPYGAFTSVAANISSLPNTYLEDVEIGTFGTPSTTSAH 300
Db 241 WEYVAFRIKTPKEGTVTIPYGAFTSVAANISSLPNTYLEDVEIGTFGTPSTTSAH 300
Qy 301 LEWMITNITLPLDRPLIS 319
Db 301 LEWMITNITLPLDRPLIS 319
RESULT 3
AAW49875
ID AAW49875 standard; protein; 319 AA.
XX
AC AAW49875;
XX
DT 17-OCT-2003 (revised)
DT 21-DEC-1998 (first entry)
XX
XX Pyrococcus VCI-7EG1 glycosidase.
XX
XX Glycosidase; VCI-7EG1; thermostable enzyme; oligosaccharide; glucose;
XX sugar; baking; textile; detergent.
XX
XX Pyrococcus furiosus; strain VCI-7EG1.
XX
XX WO9824799-A1.
XX
XX 11-JUN-1998.
XX
XX 08-DEC-1997; 97WO-US022623.
XX
XX 06-DEC-1996; 96US-0056916P.
XX
XX 10-OCT-1997; 97US-00949026.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Bylina EJ, Swanson RV, Mathur EJ, Lam DE;
XX

PH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /label= Signal_peptide
FT Protein 20..319
FT Domain /label= Mature_protein
FT Domain 28..49
FT Domain /label= Linker_moiety
FT Domain 50..319
FT Domain /label= Catalytic_domain
XX WO200196382-A2.
XX PD 20-DEC-2001.
XX PF 15-JUN-2001; 2001WO-ISO000012.
XX PR 15-JUN-2000; 2000US-00594884.
XX PA (PROK-) PROKARIA EHF.
XX PI Wicher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;
XX WPI; 2002-226850/28.
XX N-PSDB; ABA91905.
XX Novel thermostable variant cellulase which is truncated such that amino
XX terminal hydrophobic region and linker group of corresponding full length
XX enzyme is deleted, has improved catalytic properties and/or stability.
XX PS Disclosure; Page 48-49; 51pp; English.
XX CC The present sequence is that of the Pyrococcus furiosus thermostable
XX cellulase EglA, a family 12 glycosyl hydrolase. The invention provides
XX polypeptides having thermostable cellulase activity. These are variants
XX of a glycosyl hydrolase of family 12 and are truncated such that 1 or
XX more of the amino acid residues corresponding to position 1 to about 40
XX are deleted. The polypeptides are particularly derived from thermophilic
XX Rhodothermus and Pyrococcus spp. For EglA, amino acids from the N-
XX terminal hydrophobic region and/or the linking moiety are deleted.
XX Claimed polypeptides (see AAM50976-996) are also derived from
XX Rhodothermus marinus Cell12A cellulase. Isolated nucleic acids encoding
XX these thermostable cellulase polypeptides, and host cells, are also
XX claimed. The thermostable variant cellulases are useful e.g. in the wood
XX and paper pulp industries, in detergent compositions, in the textile
XX industry, to improve the feed value of animal feedstuffs, and in the food
XX industry
XX SQ Sequence 319 AA;
Query Match 100.0%; Score 1701; DB 5; Length 319;
Best Local Similarity 100.0%; Pred. No. 3.1e-150; Indels 0; Gaps 0;
Matches 319; Conservative 0; Mismatches 0;
QY 1 MSKKFVIVSILTLVQAIYVEKYHSTSEKSTNTSTPPQTTLSTTKVKIRYDDG 60
DB 1 MSKKFVIVSILTLVQAIYVEKYHSTSEKSTNTSTPPQTTLSTTKVKIRYDDG 60
QY 61 EWPAGPIDKDGNDGPEFYIEINLMNATGFAEMTYNLTSGVLHYVQOLDNIVLRDSN 120
DB 61 EWPAGPIDKDGNDGPEFYIEINLMNATGFAEMTYNLTSGVLHYVQOLDNIVLRDSN 120
QY 121 WHGYPEIFYGNKPNANVATDGPPLPSKYSNLTDFYLTISYKLEPKNGLPINFAIESW 180
DB 121 WHGYPEIFYGNKPNANVATDGPPLPSKYSNLTDFYLTISYKLEPKNGLPINFAIESW 180
QY 181 LTRAWRTTGINSDEQEMWIIYYDGLQPSGKVEIIVPIIVNGTPVNAIFEWKANIG 240
DB 181 LTRAWRTTGINSDEQEMWIIYYDGLQPSGKVEIIVPIIVNGTPVNAIFEWKANIG 240
QY 241 WEYVAFRIKTPKEGVNIPYCAFISVAANISSELPNTYELVEDVEIGTEFGTSTSAH 300
DB 241 WEYVAFRIKTPKEGVNIPYCAFISVAANISSELPNTYELVEDVEIGTEFGTSTSAH 300

QY 301 LEWWTITLTPDRPLIS 319
DB 301 LEWWTITLTPDRPLIS 319
RESULT 6
AAG79624
ID AAG79624 standard; protein; 319 AA.
XX AAG79624;
XX 07-FEB-2003 (first entry)
XX P. furiosus EglA.
XX Variant; thermostable; cellulase; Cell12A; family 12; EglA; enzyme;
XX glycosyl hydrolase; freeness; ink; coating; toner; colour; wood;
XX paper pulp; detergent; cellulose-containing textile; garment; lint;
XX fibrous crop; fruit; vegetable; grain; feed value; stability; solubility;
XX catalytic activity; cytotoxicity.
XX Pyrococcus furiosus.
XX Key Location/Qualifiers
FT Peptide 1..19
FT Peptide /label= Hydrophobic peptide
FT Peptide 28..49
FT Domain /label= Linker peptide
FT Domain 50..319
FT Domain /label= Catalytic domain
XX US2002102699-A1.
XX PD 01-AUG-2002.
XX PF 23-OCT-2001; 2001US-00003759.
XX PR 15-JUN-2000; 2000US-00594884.
XX PR 15-JUN-2001; 2001WO-ISO000012.
XX (PROK-) PROKARIA LTD.
XX Wicher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;
XX WPI; 2002-749592/81.
XX N-PSDB; ABA00497.
XX Novel isolated nucleic acid encoding a polypeptide having thermostable
XX cellulase activity, useful for producing thermostable cellulase
XX polypeptide and as probes for isolating homologous sequences.
XX Disclosure; Page 15-16; 21pp; English.
XX This sequence shows EglA, a variant family 12 glycosyl hydrolase derived
XX from P. furiosus. The amino terminal of EglA, amino acids 1-19,
XX constitute the hydrophobic domain. Residues 28-49 constitute the linker
XX moiety, with amino acids 50-319 forming the catalytic domain. EglA is
XX useful to improve freeness and to remove inks, coatings, toners and
XX colours from wood or paper pulp, in detergent compositions and to treat
XX cellulose-containing textiles and garments to improve the feel of the
XX fabric or to remove lint, in the treatment of fibrous crops, fruits and
XX vegetables or grains to improve feed values or to extract starches (e.g.
XX sugars) or other components of the crop, fruit, vegetable or grain being
XX treated. The variant glycosyl hydrolase polypeptide has improved
XX characteristics, such as increased solubility in aqueous solvents, increased
XX detergent stability, increased solubility (e.g. thermal stability,
XX catalytic activity, e.g. specific activity, catalytic rate) and/or
XX reduced cytotoxicity relative to the native or full-length thermostable
XX cellulase, but retains the substrate specificity of the native or full-
XX length cellulase
XX SQ Sequence 319 AA;

Query Match 100.0%; Score 1701; DB 5; Length 319;
Best Local Similarity 100.0%; Pred. No. 3.1e-150;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKKFVIVSILITLLVQAIYFVEKHTSDEKSTNTSSPTQTTSTTKVLKIRYPDDG 60
DB 1 MSKKKFVIVSILITLLVQAIYFVEKHTSDEKSTNTSSPTQTTSTTKVLKIRYPDDG 60

QY 61 EWPAGPIDKDGNGPEFYIEINLWNLNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSN 120
DB 61 EWPAGPIDKDGNGPEFYIEINLWNLNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSN 120

QY 121 WHGYPEIFVGNKPNWANYATDGPILPSPKVSNTDFYLTISYKLEPKNGLPIINFAIESW 180
DB 121 WHGYPEIFVGNKPNWANYATDGPILPSPKVSNTDFYLTISYKLEPKNGLPIINFAIESW 180

QY 181 LTREAWRTTGINSDEQVMIWIIYDGLQFAGSKVKEIIVPFIIVNGTPVFNATFEVWKANIG 240
DB 181 LTREAWRTTGINSDEQVMIWIIYDGLQFAGSKVKEIIVPFIIVNGTPVFNATFEVWKANIG 240

QY 241 WEYVAFRIKTIPIKEGTVTIPIYGAFISVAANISSLPNTYELYLEDVEIGTEFGTPTTSAH 300
DB 241 WEYVAFRIKTIPIKEGTVTIPIYGAFISVAANISSLPNTYELYLEDVEIGTEFGTPTTSAH 300

QY 301 LEWMITNITLTPLDRPLIS 319
DB 301 LEWMITNITLTPLDRPLIS 319

RESULT 7
ADC26955
ID ADC26955 standard; protein; 319 AA.
AC ADC26955;
XX
DT 18-DEC-2003 (first entry)
DE Pyrococcus furiosus glycosidase enzyme SEQ ID NO:60.
KW enzyme delivery matrix; thermostable enzyme; glycosidase; hydrolysis;
KW lactose; galactose; glucose; guar gum; animal feed; hydraulic fracturing;
KW oil recovery; gas recovery; biodegradable; enzyme.
XX
OS Pyrococcus furiosus.
XX
PN WO2003072717-A2.
XX
PD 04-SEP-2003.
XX
PF 21-FEB-2003; 2003WO-US0005189.
XX
PR 21-FEB-2002; 2002US-00081475.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Miller CA;
XX
PP WPI; 2003-748224/70.
DR N-PSDB; ADC26951.
XX
PT Enzyme delivery matrix used as adjuvant for animal feeds comprises
PT discrete particles each comprising grain germ that is spent of oil and
PT thermostable enzyme.
XX
PS Disclosure; SEQ ID NO 64; 103pp; English.
XX
CC The present invention describes an enzyme delivery matrix (I) comprising
CC discrete particles each comprising grain germ that is spent of oil and a
CC thermostable enzyme, where the particle readily disperses the glycosidase
CC enzyme into aqueous media. Also described: (1) a method for preparing an
CC enzyme delivery matrix; (2) isolated nucleic acid molecules encoding the
CC enzymes; (3) utilising enzymes for hydrolysing lactose to galactose and
CC glucose; (4) utilising enzymes for hydrolysing guar gum; (5) nucleic acid

CC probes; and (6) utilising enzymes or polynucleotides to generate probes.
CC (I) can be used as an adjuvant for animal feeds, and for hydraulic
CC fracturing in oil and gas recovery. (I) is biodegradable. The present
CC sequence represents a glycosidase enzyme, which is used in the
CC exemplification of the present invention.
XX
SQ Sequence 319 AA;

Query Match 100.0%; Score 1701; DB 7; Length 319;
Best Local Similarity 100.0%; Pred. No. 3.1e-150;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKKKFVIVSILITLLVQAIYFVEKHTSDEKSTNTSSPTQTTSTTKVLKIRYPDDG 60
DB 1 MSKKKFVIVSILITLLVQAIYFVEKHTSDEKSTNTSSPTQTTSTTKVLKIRYPDDG 60

QY 61 EWPAGPIDKDGNGPEFYIEINLWNLNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSN 120
DB 61 EWPAGPIDKDGNGPEFYIEINLWNLNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSN 120

QY 121 WHGYPEIFVGNKPNWANYATDGPILPSPKVSNTDFYLTISYKLEPKNGLPIINFAIESW 180
DB 121 WHGYPEIFVGNKPNWANYATDGPILPSPKVSNTDFYLTISYKLEPKNGLPIINFAIESW 180

QY 181 LTREAWRTTGINSDEQVMIWIIYDGLQFAGSKVKEIIVPFIIVNGTPVFNATFEVWKANIG 240
DB 181 LTREAWRTTGINSDEQVMIWIIYDGLQFAGSKVKEIIVPFIIVNGTPVFNATFEVWKANIG 240

QY 241 WEYVAFRIKTIPIKEGTVTIPIYGAFISVAANISSLPNTYELYLEDVEIGTEFGTPTTSAH 300
DB 241 WEYVAFRIKTIPIKEGTVTIPIYGAFISVAANISSLPNTYELYLEDVEIGTEFGTPTTSAH 300

QY 301 LEWMITNITLTPLDRPLIS 319
DB 301 LEWMITNITLTPLDRPLIS 319

RESULT 8
ABP96635
ID ABP96635 standard; protein; 300 AA.
XX
AC ABP96635;
XX
DT 02-JUN-2003 (first entry)
XX
DE Pyrococcus furiosus EGIA protein sequence SEQ ID NO:55.
XX
KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;
KW pullulanase; alpha-glucosidase; glucose isomerase; glucamylase;
KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;
KW maltodextrin; ethanol; fermentation; beverage; enzyme.
XX
OS Pyrococcus furiosus.
XX
FH Key Location/Qualifiers
FT Misc-difference 261 /note= "encoded by AAG"
XX
PN WO2003018766-A2.
XX
PD 06-MAR-2003.
XX
PF 27-AUG-2002; 2002WO-US027129.
XX
PR 27-AUG-2001; 2001US-0315281P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;
XX WPI; 2003-268420/26.
DR N-PSDB; ACC44577.
XX

PT Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.
PT alpha-amylase, useful for producing plant to produce food products having
PT improved taste or fermentable substrates for ethanol.

XX Example 26; Page 113; 158pp; English.

PS The present invention describes polynucleotides which encode processing
CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose
CC isomerase, or glucoamylase) that are optimised for expression in plants.
CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic
CC processing enzymes, which are activated under suitable conditions to act
CC upon the desired substrate. Also described are self-processing transgenic
CC plants and plant parts, e.g. grain, which express one or more of these
CC enzymes and have an altered composition that facilitates plant and grain
CC processing. Also described is a method (M) for converting starch to
CC starch-derived products in a transformed plant part (TPP), by activating
CC the starch processing enzyme contained in it. Transgenic grain is useful
CC for preparing maltodextrin. A transformed plant (TP) can be used to
CC produce food products having improved taste and to produce fermentable
CC substrates for ethanol and fermented beverages. (M) eliminates the need
CC to mill or physically disrupt the integrity of plant parts prior to
CC recovery of starch-derived products. The present sequence represents
CC ECLA, which is given in the exemplification of the present invention
XX
SQ Sequence 300 AA;

Query Match 95.1%; Score 1617; DB 6; Length 300;
Best Local Similarity 100.0%; Pred. No. 2e-142;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 IYFEKHTSDEKSTNTSTPTTSTTKVLKIRYDDGEPGAPIDKDGNGPEFYI 79
DB 1 IYFEKHTSDEKSTNTSTPTTSTTKVLKIRYDDGEPGAPIDKDGNGPEFYI 60
QY 80 EINLWNLNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSNWVHGYPEIFYGNKPNWNY 139
DB 61 EINLWNLNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSNWVHGYPEIFYGNKPNWNY 120
QY 140 ATDGPILPSKVSNLTDFTLTISYKLEPKNGLPINFAIESWLTRAWRTTGINSDEQVM 199
DB 121 ATDGPILPSKVSNLTDFTLTISYKLEPKNGLPINFAIESWLTRAWRTTGINSDEQVM 180
QY 200 IWIIYDGLQAGSKVEIWPPIIVNGTPVNATPEWKNANIGWEVAFRIKTPIKEGTVTI 259
DB 181 IWIIYDGLQAGSKVEIWPPIIVNGTPVNATPEWKNANIGWEVAFRIKTPIKEGTVTI 240
QY 260 PYGAFISVAANISSLPNTLYLEDVEIGTEFGTPSTSAHLEWWTNITLTPLDRPLIS 319
DB 241 PYGAFISVAANISSLPNTLYLEDVEIGTEFGTPSTSAHLEWWTNITLTPLDRPLIS 300

RESULT 9
AAAY23156
ID AAAY23156 standard; protein; 288 AA.

XX AAAY23156;

DT 24-AUG-1999 (first entry)

XX Dictyoglomus cellulase protein sequence.

XX Cellulose containing fabric; cellulase; fabric handle; fabric appearance;
XX pilling resistance.

OS Dictyoglomus sp.

FN WO9932708-A1.

XX 01-JUL-1999.

XX 17-DEC-1998; 98WO-US026798.

XX 19-DEC-1997; 97US-0068274P.

XX (NOVO) NOVO NORDISK BIOCHEM NORTH AMERICA.
PA Liu J, Condon B;
PI WPI; 1999-395422/33.
DR Treatment of cellulose containing fabric using cellulase, achieves better
XX handling, appearance and pilling resistance.
XX Claim 7; Page 26; 31pp; English.
XX The specification describes a method for treating a cellulose containing
CC fabric, comprising contacting the fabric with an aqueous bulk solution
CC containing a cellulase, and subjecting the contacted fabric to high
CC temperature. The method of treatment is useful for treating cellulose-
CC containing fabrics. The method of the invention achieves better fabric
CC handle, appearance and pilling resistance. The quality of the cellulosic
CC fabric is enhanced and allows uniform action by the cellulase. The
CC present sequence represents a cellulase which is used in the method of
CC the invention
XX
SQ Sequence 288 AA;

Query Match 28.6%; Score 486.5; DB 2; Length 288;
Best Local Similarity 32.2%; Pred. No. 7.7e-37;
Matches 104; Conservative 58; Mismatches 106; Indels 55; Gaps 5;
QY 4 KKFVIVSILTLILVQAIY-----FVEKYHTSDEKSTNTSTPTTSTTKVLKI 54
DB 2 KKSLLSLILILLITLSPSQPKYKDAFILKAPSSGDVTTKNLEPLT----- 47
QY 55 RYDDGEPGAPIDKDGNGPEFYIENLWNLNATGFAEMTYNLTSGVLHYVQQLDNIV 114
DB 48 -----DELNFNLIANYEGNTWAFYKEEDTVEYYADIKNIV 83
QY 115 LRDRSNWVHGYPEIFYGNKPNWNYATDGPILPSKVSNLTDFTLTISYKLEPKNGLPIN 174
DB 84 LKDKSNWVHGYPEIFYGYKPNWAGHNSIEKLALPKVSEFPDVLNKNIWEKKNLPIN 143
QY 175 FAIESWLTRAWRTTGINSDEQVMIIYYDGLQAGSKVEIWPPIIVNGTPVNATPEV 234
DB 144 FAMEWTIKPEYQKT-VTSGDIEMWVLYANRLSPAGKGVKIPILNNGNQDIIWEV 202
QY 235 WZANIGWEVAFRIKTPIKEGTVTIPYGAFIS-----VAANISSL--PNYTELYLEDVEI 287
DB 203 YLSPMSWVYVAYKXENILQGVKIPINEFLKHLRTILANNPSRITPEKFDQMYVTWVEI 262
QY 288 GTEFGTPSTSAHLEWWTNITLT 310
DB 263 GTEFGDPVTTEAKFGWTFSEFDI 285

RESULT 10
AAW57777

ID AAW57777 standard; protein; 297 AA.

XX AAW57777;

DT 17-OCT-2003 (revised)

DT 26-OCT-1998 (first entry)

XX Recombinant endo-1,4-beta-glucanase.

XX Endoglucanase; cellulase; thermostable enzyme; textile; biopolishing;
XX stone-washing; saccharification; feedstuff; coffee.

OS Bacillus sp.

OS Dictyoglomus; sp.

OS Chimeric.

XX Key Location/Qualifiers

FT Peptide 1. .27

Query Match 28.2%; Score 480; DB 2; Length 297;
Best Local Similarity 32.1%; Pred. No. 3.3e-36;
Matches 105; Conservative 63; Mismatches 109; Indels 50; Gaps 7;
QY 1 MSKKKFFVIVSLTLLVQALFYFVEKYHTSBDKSTSTSTP-----POTTLSTTK 50
DB 1 MKQQRLLYARLLTLTF--ALIFLLPH-----SAAAAQTPKYKDAFILKAPSSGDVTK 51
QY 51 VLKIRYDDGWCAPIDKDGNGPEFYIEINLWNLNATGFAEMTNLTSGVLHVYQOL 110
DB 52 NLPT-----LELNFNFIANYEGNTWMAFYKEEDVEYADI 88
QY 111 DNVLRDRSNVWHGYPEIFYGNKPNWNYATDGPILPSPKSVNLTDPLYISYKLPKNG 170
DB 89 KNIVLKDKNSVWHGYPEIFYGYKDWAGHGSIEKLALPKVSEFPDVLFLNLYNVEKN 148
QY 171 LPINFAESLWLTREAWRTTGINSDEQEVMIWYDGLQAPGSKVKEIVVPIVNGTPVNA 230
DB 149 LPINFAMETWTKEPYQKT-VTSGDIEVMVWLVANRUSPAGRKVGKIPFIILNGKQDI 207
QY 231 TFEVVKANIGWEYVAFRIKTIKEGTVTIPIYGAFIS-----VAANISSL--PNYTELYLE 283
DB 208 IWEYVLSFMSWDYVAYKSKENILQGVKIPINEFLKHLRTILANNPSRIITPEKFDQMYVT 267
QY 284 DVEIGTEFGTPSTSAHLEWITNITL 310
DB 268 VWEIGTEFGDPYTTTEAKFGWTFNSFDI 294

RESULT 12
AAV17865
ID AAV17865 standard; protein; 332 AA.
AC AAV17865;
XX
XX
DT 17-AUG-1999 (first entry)
XX
DE Sulfolobus solfataricus endo-beta-1,4-glucanase Cella.
XX
XX Sulfolobus solfataricus; endoglucanase; Cella; CelB;
KW endo-beta-1,4-glucanase.
XX
XX Sulfolobus solfataricus.
XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label= signal
FT Protein 24..332
FT /label= Cella
XX
XX DK9900097-A.
PN
XX
PD 12-JAN-1999.
XX
XX 12-JAN-1999; 99DK-000000097.
PF
XX
PR 12-JAN-1999; 99DK-000000097.
XX
XX (NOVO) NOVO-NORDISK AS.
PA
XX
XX WPI; 1999-279376/24.
DR N-PSDB; AAX80187.
XX
XX Novel endonuclease - produced by Sulfolobus solfataricus.
PT
XX
XX Claim 1; Page 27-28; 36pp; English.
PS
XX
CC The present sequence represents a Sulfolobus solfataricus endoglucanase,
CC specifically endo-beta-1,4-glucanase, designated Cella. The new endo-beta-
CC 1,4-glucanase is produced by Sulfolobus solfataricus ATCC 35092
XX
XX Sequence 332 AA;
SQ

Query Match 25.8%; Score 439.5; DB 2; Length 332;
Best Local Similarity 33.4%; Pred. No. 2.3e-32;
Matches 117; Conservative 57; Mismatches 121; Indels 55; Gaps 17;
QY 1 MSKKKFFVIVSLTLLV-----QAIYFVEKYHTSBDKSTSN-----TGSTP 41
DB 1 MNKLVIIIVPIIVGVIGGAIYL---HHQSNVKTSSITVTNETTILMSITINVP 57
QY 42 POTTLSTTKVLKIRYDDGWCAPIDKDGNGPEFYIEINLWNLNATGFAEMTNLTSL 101
DB 58 TTVPTTSSIQLIYVTVSSASSPTFVYLNNTSVSFYLEVNMNNAKTNGNYTVWFENPLT 117
QY 102 GVLHVQOLDNIVLRDRSNVWHGYPEIFYGNKPNWNYATDGPILPSPKSVNLT----DF 157
DB 118 RTLSVSFNLTVQ--NPLQWNTNGYPEIYVGRKPDWTSYA--GNI-FPMRIGNMTTFMVSE 171
QY 158 YLTISYKLEPKNGLPINF--AIESWLTRAWR---TTGINSDEQEVMIWYDGLQAPGS 212
DB 172 YINLT-KLDPS-----INFDIASDAWIVRPOAFSPGTAPGNGDIEIMVWLFESQLQAPG 226
QY 213 KYKEIVVPIVNGTPVNAATFEWK-ANI---GWEYVAFRIKTP-----IKEGTVTIPIYCAP 264
DB 227 QVGEVVIPIYINHTLVNATFQWKMKNVPGWCGWBYIAFR---PDGKVTNGYVAYEPNLF 283
QY 265 ISVAANISSLPNYTELYLDEVEIGTGT-PSTTSAHLEWITNITLPL 313
DB 284 IKALNNFASY-NITNYLTDWBEFTEWGTMTSNGTAIFYSWTISNFYETLL 332
RESULT 13
AAV17866
ID AAV17866 standard; protein; 332 AA.
AC AAV17866;
XX
XX
DT 17-AUG-1999 (first entry)
XX
DE Sulfolobus solfataricus endo-beta-1,4-glucanase CelB.
XX
XX Sulfolobus solfataricus; endoglucanase; Cella; CelB;
KW endo-beta-1,4-glucanase.
XX
XX Sulfolobus solfataricus.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= signal
FT Protein 22..332
FT /label= CelB
XX
XX DK9900097-A.
PN
XX
PD 12-JAN-1999.
XX
XX 12-JAN-1999; 99DK-000000097.
PF
XX
PR 12-JAN-1999; 99DK-000000097.
XX
XX (NOVO) NOVO-NORDISK AS.
PA
XX
XX WPI; 1999-279376/24.
DR N-PSDB; AAX80188.
XX
XX Novel endonuclease - produced by Sulfolobus solfataricus.
PT
XX
XX Claim 1; Page 31-32; 36pp; English.
PS
XX
CC The present sequence represents a Sulfolobus solfataricus endoglucanase,
CC specifically endo-beta-1,4-glucanase, designated CelB. The new endo-beta-
CC 1,4-glucanase is produced by Sulfolobus solfataricus ATCC 35092
XX
XX Sequence 332 AA;
SQ

Query Match 25.3%; Score 430; DB 2; Length 332;
Best Local Similarity 32.2%; Pred. No. 1.8e-31;
Matches 111; Conservative 60; Mismatches 124; Indels 50; Gaps 15;

QY 4 KKFVIVSILTLV-----QALFYVEKHTSEDKSTSTSTP-----PQTLL 46
DB 3 KLYIVLPIVIAIAGVNGGIIYLHQOGLSVKPVTTTTFSTTSTTTTNAITTTVT 62
QY 47 STTKVLKIRYPPDGGWPGAPIDKDGNGPFFYIEINLWLNILNATGPAEMTYNLTSGVLHY 106
DB 63 SITSYNQLIIVTSASASPTPYVNNSTIPSYFLEVNNWNAKNNGYNTWVFNPLARTLSV 122
QY 107 VQQLDNIVLRDRSNWHVGYEIPFYGNKPNWANYATDGPFLPSKVNLT----DFYLTIS 162
DB 123 SFNLTQV---KPLEWTNGYPEIYVGRKPDYTA--GNI--FPMRIGNMTPEMVSFYINLT 176
QY 163 YKLEPKNGLEINF--ATESLITREAWR---TTGINSDEQEVMTWYYDGLQPAGSKVKEI 217
DB 177 -KLDPS-----INFDIASDAWIVRQIAFSPGTAAGNGDIEIMVLFSONLQPAGEQVGKV 231
QY 218 VVPIVNGTVPNATFFEVWKANI-----GWYVAFRIKTP-----IKEGTVTIPYGAFISVAA 269
DB 232 VVPIYINHTLVNATFQWEMKSPWGGWEYIAFR---PDGKVTNGVVSVEPNLFIKALS 288
QY 270 NISSLPNYTELYLEDVEIGTEPGT-PSTTSAHLEWITNITLPL 313
DB 289 NFTSY-NITNYXLTDFEFGTEWGTMTSNGTAYFSWTVSFSETLL 332

RESULT 14

AAW34985

ID AAW34985 standard; protein; 841 AA.

AC AAW34985;

XX 27-AUG-2003 (revised)

DT 21-MAY-1998 (first entry)

XX Archaebacterial thermostable endoglucanase.

XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;

XX beta-1,4-glycosidic bond; hydrolysis; saccharification;

XX archael bacterium; thermostable enzyme; thermophilic.

XX archaeon.

XX Unidentified.

XX WO9744361-A1.

XX 27-NOV-1997.

XX 22-MAY-1997; 97WO-US008793.

XX 22-MAY-1996; 96US-00651572.

XX (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX Lam DE, Mathur EJ;

XX WPI; 1998-018435/02.

XX N-PSDB; AAT94193.

XX Endoglucanase(s), preferably from archael bacterium, AEPII 1a - useful

XX to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic

XX bonds in cellulose.

XX Claim 1; Fig 1A; 164pp; English.

XX This protein comprises a thermostable 60.9 kDa endoglucanase of archael

XX bacterial strain AEPIIa, a thermophilic isolate of a marine hydrothermal

XX vent. The endoglucanase is capable of degrading carboxymethylcellulose

XX and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It can be

XX produced from native cells or from recombinant host cells, especially

CC prokaryotic host cells transformed with a plasmid or virus-derived vector
CC including the endoglucanase DNA (see AAT94193). 23 Other homologous
CC endoglucanases (see AAW34986-W35008) are also claimed. The endoglucanases
CC can be used to degrade cellulose for the conversion of plant biomass into
CC fuels and chemicals, for use in detergents, textiles, animal feed, waste
CC treatment, and in the fruit juice and brewing industries for the
CC clarification and extraction of juices. (Updated on 27-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 841 AA;

Query Match 19.3%; Score 328; DB 2; Length 841;

Best Local Similarity 67.4%; Pred. No. 2.3e-21;

Matches 58; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 28 TSEKSTNTSTPTQTLLSTTKVLKIRYPPDGGWPGAPIDKDGNGPFFYIEINLWLNIL 87

DB 756 TTTQSTTTTTTSPPTTTAPADVKIKRYPPDGGWPGAPIDKDGNGPFFYIEINLWLNIL 815

QY 88 NATGPAEMTYNLTSGVLHYVQQLDNI 113

DB 816 SAESYAEMTYNLTSGVLHYVQQLDNI 841

RESULT 15

AAW50977

ID AAW50977 standard; protein; 243 AA.

XX AAW50977;

XX 15-MAY-2002 (first entry)

DT Truncated thermostable cellulase Cell12A (aal9-261).

XX Cellulase; Cell12A; thermostable; enzyme; mutant; mutein.

XX Rhodothermus marinus.

XX Synthetic.

XX WO200196382-A2.

XX 20-DEC-2001.

XX 15-JUN-2001; 2001WO-IS0000012.

XX 15-JUN-2000; 2000US-00594884.

XX (PROK-) PROKARIA EHF.

XX Wicher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;

XX WPI; 2002-226850/28.

XX N-PSDB; ABA91885.

XX Novel thermostable variant cellulase which is truncated such that amino

XX terminal hydrophobic region and linker group of corresponding full length

XX enzyme is deleted, has improved catalytic properties and/or stability.

XX Claim 15; Page; 51pp; English.

XX The present sequence corresponds to amino acids 19-261 of Rhodothermus

XX marinus thermostable cellulase Cell12A. This claimed truncated variant of

XX Cell12A therefore lacks the N-terminal hydrophobic region and 1 of the

XX amino acids of the linker moiety of the native enzyme (see also

XX AAW50975). 21 Truncated variants of Cell12A (see AAW50976-996), and

XX isolated nucleic acids encoding them (see ABA91884-904), are claimed. The

XX variants have a half-life of at least 3.5 hours at 90 degrees C. They are

XX more soluble than the native enzyme, and have a specific activity that is

XX at least 2 times greater than that of the native enzyme. Fusion proteins

XX comprising a variant thermostable cellulase and a fusion partner are also

XX claimed, as well as host cells and methods of production. The

XX thermostable variant cellulases are useful e.g. in the wood and paper

XX pulp industries, in detergent compositions, in the textile industry, for

CC improving the feed value of animal feedstuffs, and in the food industry.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from the Cell12A sequence given in the Sequence Listing (see
 CC AAM50975)

XX
 SQ Sequence 243 AA;

Query Match	9.8%;	Score 166.5;	DB 5;	Length 243;
Best Local Similarity	25.6%;	Pred. No. 4.8e-07;		
Matches	66;	Conservative 35;	Mismatches 108;	Indels 39; Gaps 11;

QY	59	DGEWPGAPIDKDGNGPEFYIEI-----NLWNLNATCFAMTYNLATSGVLHYV	107
DB	1	DWLFPDGDNGKEPEPEPTVELCGRWDRDARDVAGGRYRVINNVWGAETAOCIEVGL----	56
QY	108	QQLDNIVL----RDRSNWVHGYPEIFYGNKPNWANYATDGPILPSKVSNLTDFTYLTISY	163
DB	57	-ETGNFTITRADHDNGNNVAAYPAYFGCH-WGACTSNG--LPRRVQELSD--VRTSW	109
QY	164	KLEPANGLPINFAIESWLITREAWTTTGINSDEQEVMIYVD-GLOPAGSKVKEIVVPIL	222
DB	110	TLTPTTGRWNAAYDIWFSPVTSNGY-SGGAEIMLWLNWNGVMPGGRVATVEL---	165
QY	223	VNGTPVNATFEVWKANIGWEYVAFRIKTIKEGTVTIPYGAFTISVAANISSEPNYTELYL	282
DB	166	-----AGATWVWYADWDWNYIARVTRTP-TTSVSELDLKAFFIDDAVARGYI--RPEWYL	217
QY	283	EDVEIGTE	290
DB	218	HAVETGFE	225

Search completed: June 29, 2004, 20:11:34
 Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2004, 14:39:56 ; Search time 6088 Seconds
(without alignments)
6834.644 Million cell updates/sec

Title: US-09-914-543-45
Perfect score: 960
Sequence: 1 atgagcaagaaagtctgt.....atagacctcttatttcttaa 960

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1:	gb_ba.*	GenEmbl.*
2:	gb_hcg.*	
3:	gb_in.*	
4:	gb_on.*	
5:	gb_ov.*	
6:	gb_pat.*	
7:	gb_ph.*	
8:	gb_pl.*	
9:	gb_pr.*	
10:	gb_ro.*	
11:	gb_sts.*	
12:	gb_sy.*	
13:	gb_un.*	
14:	gb_vi.*	
15:	em_ba.*	
16:	em_fun.*	
17:	em_hum.*	
18:	em_in.*	
19:	em_mu.*	
20:	em_or.*	
21:	em_ov.*	
22:	em_pat.*	
23:	em_ph.*	
24:	em_pl.*	
25:	em_ro.*	
26:	em_sts.*	
27:	em_un.*	
28:	em_vi.*	
29:	em_hcg_hum.*	
30:	em_hcg_inv.*	
31:	em_hcg_other.*	
32:	em_hcg_mus.*	
33:	em_hcg_pln.*	
34:	em_hcg_rod.*	
35:	em_hcg_mam.*	
36:	em_hcg_vrt.*	
37:	em_sy.*	
38:	em_hgto_hum.*	
39:	em_hgto_mus.*	
40:	em_hgto_other.*	
41:	em_hgto_mus.*	

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	960	100.0	960	6	AR205123	Sequence
2	960	100.0	960	6	BD056827	A thermos
3	960	100.0	960	6	BD064594	Glycosida
4	960	100.0	11753	1	AE010200	Pyrococcu
5	958.4	99.8	1134	1	AF181032	Sequence
6	958.4	99.8	1134	6	AX339683	Sequence
7	114.2	11.9	867	6	BD023645	Novel end
8	89.2	9.3	2803	1	TN93354	Thermotoga
9	84.2	8.8	2800	1	TMCELAB	Z69341 T.maritima
10	84.2	8.8	20853	1	AE001800	Thermotog
11	82	8.5	4547	1	TN286103	T.neapolita
12	53.8	5.6	7218	6	AE055393	Sequence
13	47.4	4.9	2000	6	AX655393	Sequence
14	47.2	4.9	11595	1	AE006803	Sulfolobu
15	46.4	4.8	37037	3	CEP17C11	Caenorhabdi
16	46.2	4.8	128683	2	AC135793	Oryza sat
17	44.2	4.6	13962	1	AE006749	Sulfolobu
18	44	4.6	128683	2	AC135793	Oryza sat
19	43.6	4.5	53561	1	AE000790	Borrelia
20	42.8	4.5	193018	2	EX324226	Danio rer
21	42.4	4.4	1141	6	AX083744	Sequence
22	42.2	4.4	184730	2	AC146676	Papio anu
23	41	4.3	134971	2	AC116367	Oryza sat
24	40.6	4.2	166885	5	EX005129	Zebrafish
25	40.6	4.2	211342	2	EX511164	Danio rer
26	40.2	4.2	144781	9	HSJ585114	Human DNA
27	39.6	4.1	129332	9	AL590028	Human DNA
28	39.6	4.1	149646	2	AL163536	Homo sapi
29	39.6	4.1	160708	2	AC068406	Homo sapi
30	39.6	4.1	163603	2	AC026755	Homo sapi
31	39.6	4.1	197850	2	AC027690	Homo sapi
32	39.6	4.1	201823	9	HSJ210K12	Human DNA
33	39.4	4.1	132202	9	AC025218	Homo sapi
34	39.4	4.1	184772	2	AC114607	Mus muscu
35	39.4	4.1	188627	10	AL844221	Mouse DNA
36	39.4	4.1	194042	2	EX890633	Mus muscu
37	39.4	4.1	265693	2	AC122319	Mus muscu
38	39.4	4.1	274954	2	AC131556	Mus muscu
39	39.2	4.1	239840	2	AC094367	Rattus no
40	39.2	4.1	253095	2	AC130143	Rattus no
41	38.8	4.0	91202	2	AC007609	Homo sapi
42	38.4	4.0	11892	10	AL929538	Mouse DNA
43	38.4	4.0	154864	9	AC143356	Pan trogl
44	38.4	4.0	256680	2	AC098136	Rattus no
45	38.2	4.0	16076	3	AB055862	Procambur

ALIGNMENTS

RESULT 1	AR205123	Sequence 60 from patent US 6368844.	DNA	linear	PAT 20-JUN-2002
LOCUS	AR205123	Sequence 60 from patent US 6368844.			
DEFINITION	AR205123	Sequence 60 from patent US 6368844.			
ACCESSION	AR205123	Sequence 60 from patent US 6368844.			
VERSION	AR205123.1	GI:21502626			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 960)				
AUTHORS	Bylina E.J.				
TITLE	Glycosidase enzymes				
JOURNAL	Patent: US 6368844-A 60 09-APR-2002;				
FEATURES	Location/Qualifiers				


```
Db 541 TTAACGAGAGAGCGTTGGAGAACACACAGGAATTAACAGCGATGACGAAGTAATGATA 600
Qy 601 TGGATTACTATGACGATTACACCGCTGGCTCCAAAGTTAAGGAGATTGAGTCCCA 660
Db 601 TGGATTACTATGACGATTACACCGCTGGCTCCAAAGTTAAGGAGATTGAGTCCCA 660
Qy 661 ATAATAGTTAAACGGAACACACAGTAATGCTACATTTGAAGTATGGAAGCAAACTGGT 720
Db 661 ATAATAGTTAAACGGAACACACAGTAATGCTACATTTGAAGTATGGAAGCAAACTGGT 720
Qy 721 TGGAGTATGCTGCATTTAGATTAAGACACCCCAATCAAGAGGAGACAGTGACAATCCA 780
Db 721 TGGAGTATGCTGCATTTAGATTAAGACACCCCAATCAAGAGGAGACAGTGACAATCCA 780
Qy 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTTACAGAACTT 840
Db 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTTACAGAACTT 840
Qy 841 TACTTAGAGACGTTGGAGATTGNACTGAGTTTGGAAAGCGCAAGCACTACTCGGCCAC 900
Db 841 TACTTAGAGACGTTGGAGATTGNACTGAGTTTGGAAAGCGCAAGCACTACTCGGCCAC 900
Qy 901 CTAGAGTGGTGGATACAAACATACACTAACTAATCTCTAGATAGACCTCTTATTCTTAA 960
Db 901 CTAGAGTGGTGGATACAAACATACACTAATCTCTAGATAGACCTCTTATTCTTAA 960

RESULT 3
BD064594
LOCUS BD064594 960 bp DNA linear PAT 27-AUG-2002
DEFINITION Glycosidase enzymes.
ACCESSION BD064594
VERSION BD064594.1 GI:22610197
KEYWORDS JP 2001505440-A/46.
SOURCE Vaccinia virus
ORGANISM Vaccinia virus
Virus; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae; Orthopoxvirus.
REFERENCE 1 (bases 1 to 960)
AUTHORS Bylina,E.J., Swanson,R.V., Mathur,E.J. and Lam,D.E.
TITLE Glycosidase enzymes
JOURNAL Patent: JP 2001505440-A 46 24-APR-2001;
DIVERSA CORP
COMMENT PN JP 2001505440-A/46
PD 24-APR-2001
PF 08-DEC-1997 JP 1998525887
PR 06-DEC-1996 US 60/056916
PI EDWARD J BYLINA, RONALD V SWANSON, ERIC J MATHUR, DAVID E LAM PC
C07H21/04, C12N1/20, C12N1/14, C12N5/00, C12N9/38, C12N9/42, C08B30/
04
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1..960
/organism="Vaccinia virus"
/mol_type="genomic DNA"
/db_xref="taxon:10245"
ORIGIN
Query Match 100.0%; Score 960; DB 6; Length 960;
Best Local Similarity 100.0%; Pred. No. 3.7e-255;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGAGCAGAAAAGTTGTCATCGTATCTATCTTAATCTTAATCTTTAGTACAGGCAATA 60
Db 1 ATGAGCAGAAAAGTTGTCATCGTATCTATCTTAATCTTAATCTTTAGTACAGGCAATA 60
Qy 61 TATTTTGTAGAAAAGTATCATACCTCTGAGGACAGTCAACTTCAAAATACCTCATCTACA 120
Db 61 TATTTTGTAGAAAAGTATCATACCTCTGAGGACAGTCAACTTCAAAATACCTCATCTACA 120
```

```
Qy 121 CCACCCCAAAACACACTTTCCACTACCAAGGTTCTCAAGATTAGTACCCTTGATGACGGT 180
Db 121 CCACCCCAAAACACACTTTCCACTACCAAGGTTCTCAAGATTAGTACCCTTGATGACGGT 180
Qy 181 GAGTGGCCAGGAGCTCCCTATTGATAAGGATGGTATGGAAACCCAGAAATCTTACATTGAA 240
Db 181 GAGTGGCCAGGAGCTCCCTATTGATAAGGATGGTATGGAAACCCAGAAATCTTACATTGAA 240
Qy 241 ATAAACCTTATGGAAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTACAAATTAACC 300
Db 241 ATAAACCTTATGGAAACATTTCTTAATGCTACTGATTTGCTGAGATGACGTACAAATTAACC 300
Qy 301 ACGCGGCTCTTCACTAGCTCCCAACACTTGACACACTTGTCTTGAGGATAGAAATTAAT 360
Db 301 ACGCGGCTCTTCACTAGCTCCCAACACTTGACACACTTGTCTTGAGGATAGAAATTAAT 360
Qy 361 TGGGTGCTATGGATACCCCGAAATATTCTATGGAAACAAGCCATGGAATGCAAACTACGCA 420
Db 361 TGGGTGCTATGGATACCCCGAAATATTCTATGGAAACAAGCCATGGAATGCAAACTACGCA 420
Qy 421 ACTGATGGCCCAATACCATACCCAGTAAAGTTTCAAACTTAACAGACTTCTATCTTACA 480
Db 421 ACTGATGGCCCAATACCATACCCAGTAAAGTTTCAAACTTAACAGACTTCTATCTTACA 480
Qy 481 ATCTCTATAAACTTTGAGCCCAAGAACCGGCTGCCAATTAACTTCGCAATAGAATCTGG 540
Db 481 ATCTCTATAAACTTTGAGCCCAAGAACCGGCTGCCAATTAACTTCGCAATAGAATCTGG 540
Qy 541 TTAACGAGAGAGCTTTGGAGAACAAACAGAAATTAACAGCGATGACGAAGTAATGATA 600
Db 541 TTAACGAGAGAGCTTTGGAGAACAAACAGAAATTAACAGCGATGACGAAGTAATGATA 600
Qy 601 TGGATTACTATGACGATTACACCGCTGGCTCCAAAGTTAAGGAGATTGAGTCCCA 660
Db 601 TGGATTACTATGACGATTACACCGCTGGCTCCAAAGTTAAGGAGATTGAGTCCCA 660
Qy 661 ATAATAGTTAAACGGAACACACAGTAATGCTACATTTGAAGTATGGAAGCAAACTGGT 720
Db 661 ATAATAGTTAAACGGAACACACAGTAATGCTACATTTGAAGTATGGAAGCAAACTGGT 720
Qy 721 TGGAGTATGTTGCAATTTAGATAAAGACCCCAATCAAGAGGAGACAGTGACAATCCA 780
Db 721 TGGAGTATGTTGCAATTTAGATAAAGACCCCAATCAAGAGGAGACAGTGACAATCCA 780
Qy 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTTACAGAACTT 840
Db 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTCAAGCTTTACCAATTTACAGAACTT 840
Qy 841 TACTTAGAGACGTTGGAGATTGGAATCTGATTTGGAACGCCAAGCACTACTCGGCCAC 900
Db 841 TACTTAGAGACGTTGGAGATTGGAATCTGATTTGGAACGCCAAGCACTACTCGGCCAC 900
Qy 901 CTAGAGTGGTGGATCACAACATACACTAATCTCTAGATAGACCTCTTATTCTTAA 960
Db 901 CTAGAGTGGTGGATCACAACATACACTAATCTCTAGATAGACCTCTTATTCTTAA 960

RESULT 4
AE010200/c 11753 bp DNA linear BCT 25-FEB-2002
LOCUS Pyrococcus furiosus DSM 3638, section 75 of 173 of the complete genome.
DEFINITION
ACCESSION AE010200 AE009950
VERSION AE010200.1 GI:18892881
KEYWORDS Pyrococcus furiosus DSM 3638
SOURCE Pyrococcus furiosus DSM 3638
ORGANISM Pyrococcus furiosus DSM 3638
REFERENCE 1 (bases 1 to 11753)
AUTHORS Maeder,D.L., Weiss,R.B., Dunn,D.M., Cherry,J.L., Gonzalez,J.M., DiRuggiero,J. and Robb,F.T.
Divergence of the hyperthermophilic archaea Pyrococcus furiosus and
```

```

P. horikoshii inferred from complete genomic sequences
Genetics 152 (4), 1299-1305 (1999)
MEDLINE 99359404
PUBMED 10430560
REFERENCE 2 (bases 1 to 11753)
AUTHORS Robb,F.T., Maeder,D.L., Brown,J.R., DiRuggiero,J., Stump,M.D.,
Yeh,R.K., Weiss,R.B. and Dunn,D.M.
TITLE Genomic sequence of hyperthermophile, Pyrococcus furiosus:
implications for physiology and enzymology
Meth. Enzymol. 330, 134-157 (2001)
MEDLINE 21079003
PUBMED 11210495
REFERENCE 3 (bases 1 to 11753)
AUTHORS Weiss,R.B., Dunn,D.M., Robb,F.T. and Brown,J.R.
TITLE The complete sequence of the Pyrococcus furiosus genome
JOURNAL Unpublished
AUTHORS Weiss,R.B.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-2002) Human Genetics, University of Utah, 20
South 2030 East, Salt Lake City, UT 84112, USA
FEATURES
    source
        1..11753
        /organism="Pyrococcus furiosus DSM 3638"
        /mol_type="genomic DNA"
        /strain="DSM 3638"
        /db_xref="taxon:186497"
        166..1242
        /gene="PF0849"
        166..1242
        /gene="PF0849"
        166..1242
        /note="Function Code: 16.1 Conserved Hypothetical"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAL80973.1"
        /db_xref="GI:18892882"
        /translation="MYELTDFDKLRKITYKVLGDVDERDLNVPSSKAGPCGNCL
        FLYLQNPENIYVAVSHDINDPTLEERIRYARKHYDLWIRVDTSGNVKFDENIK
        SCYAGLDEIOISVHTTKDVRKLMNRHAGKIDLLPLVAKHFRTIADIILTPGFN
        VDDIGEIEDDSMGHEVRLFPVGVTKYRFEIRPLTKELSVMKEVALEKDELGI
        KVVPIRPIFLLGFTTGLLEPNLEPEFFYITGELAYEMKRLPRLIKVAVKNEF
        FCGNIGAGLTGRDVREVERLPEVDVFGILLPELMFYGDMLDGNRRQDLFSKILI
        EKVIVTALTEPTEIPKVIKISL"
        1235..1289
        /note="small nucleolar RNA, posttranscriptional
        modification of rRNA"
        1265..2380
        /gene="PF0850"
        1265..2380
        /gene="PF0850"
        /note="Function Code: 16.1 Conserved Hypothetical"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAL80974.1"
        /db_xref="GI:18892883"
        /translation="MIRGLAEGRMKETSYSYVVGELPKGCQYCVRGKULVFTG
        VCPNCPYCLSPWRKDVAYANERPIRKIEDIEEAKIQEARGAGVTGGDPLARLN
        TVEYIRLKEEFQKKFHLYTIGILADKRALSMYDAGLDEIRFPDLPFSSKFLN
        REIENLNAPFDMDVGVEVPVAPFGDRIKWFAEILDYFGKFLNLELEYSETNLS
        NILNRGPKLAGENFAIKGSLGLEILEWKENKTSLYHLCTAKLKDAVOLNRLLR
        MAKVAPYMEITEDGTLRGIAVEDLEELYKFLVEEAEPVKENLYINWEKRIEMP
        VEVAEELAELEGVDFYVEEPTWDRIEVERIPLN"
        2459..2860
        /gene="PF0851"
        2459..2860
        /gene="PF0851"
        /note="Function Code: 10 Regulatory Functions: REGULATORY
        FUNCTIONS"
        /codon_start=1
        /transl_table=11
        /product="iron-dependent repressor"

/protein_id="AAL80975.1"
/db_xref="GI:18892884"
/translation="MYIQKNGVIRVDIAMMRVKPPTVVEALKKLRDKGVKVEE
HEHILLTKEGVAKTKYSKQLLTFFINILGIPPEIAERDACCQFHYVSEVTVHRI
REFISYIQECPYALKQFLKKVREKDAVAK"
complement(2842..4170)
/gene="PF0852"
complement(2842..4170)
/gene="PF0852"
/note="Function Code: 16.1 Conserved Hypothetical"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL80976.1"
/db_xref="GI:18892885"
/translation="MGNWIERFEPEKYGTDLKTVEVLGITTFMTMAIIFVNPILLSN
AMGPEAFDSLVAATSLAAGLTILMGIYAKPFALPGMGLNPAFTVAPKYGKVA
LAAVEGEGIIIFISITKRSATTHAIPLSQYAGAGIGLFTFGLNNVGLLTALT
GDNGLLIKTSLNASVLARPEGLFLFGLTALILSLIKGALLISILATSIIGMITG
VAPRPEGVSMPTLSYTFQLDLKLNAGALGAVFAPFMVDFDTLTGTGLTGLSAKAG
ELTKEGKIPDAEKVLLTDAIGTGTGAILGTSTVTYTESAAGIBEGRTGPTAVITGL
LELIIGLSIAPVAKAIPAPATAPALVINGYWSAKFVEDPSDEALPAFLVLTIP
FYISADGIGVGFISYTLKVFTRGNKEVHPMLTALAVVFIGYFAYLATA"
complement(4316..5113)
/gene="PF0853"
complement(4316..5113)
/gene="PF0853"
/note="Function Code: 9.5 Purines, Pyrimidines,
Nucleosides and Nucleotides: Salvage of nucleosides and
nucleotides"
/codon_start=1
/transl_table=11
/product="5'-methythioadenosine phosphorylase II"
/protein_id="AAL80977.1"
/db_xref="GI:18892886"
/translation="MPRIATVGGSGVYDFPAENKREETVTKPYGEVKITVGWGDDEV
AFLRKHGHSIPHKINRYANIALWELHGVRIIATSAVGSMMPEKPGDFVILDQI
IDFTVSRPRTFYDGEESPHERRFVHVDFTFPCPEIRKALITATARNLGLDHPRGTY
VCTGEPRTAAEIRAYRILGCVVGTQCPEAILARELEMCYATVAIVTNYAAGMSG
KKLTSEVSEVMKQKSDIVKLIILAAIPLIKERRCGCKALKGATG"
5176..5253
/product="tRNA-Arg"
/note="tRNA Arg anticodon CCG"
complement(5400..6359)
/gene="PF0854"
complement(5400..6359)
/gene="PF0854"
/note="Function Code: 5.1 Central Intermediary Metabolism:
Degradation of polysaccharides"
/codon_start=1
/transl_table=11
/product="endo-1,4-beta-glucanase b"
/protein_id="AAL80978.1"
/db_xref="GI:18892887"
/translation="MSKKKFIVSITLITLLVQAIYFVEKHTSDKSTNTSTPPQT
TLSTTKVLKIRVPDDGWPAPIDKDGDNPEFYIEINLWILNATGPAETMYNLTSQ
VLHYVQDLNIVLRDSRWVHGYEIEFYKGNKNVANYATDGPILPSKVSNLTDYLT
ISYKPKNGCDIPNFAIESWLTREAWRTGINSDEQEWMIWYIDGLQPGASKVEIV
VPILVNGPKNATEVWKNAGIWEYFAPRIKPIKEGTVTPIYDGAIFSAVNAISLEN
YTEIYLEVEITGTFGTSTTSAAHLEWITNITILTPDRPLIS"
complement(6356..7246)
/gene="PF0855"
complement(6356..7246)
/gene="PF0855"
/note="Function Code: 16.1 Conserved Hypothetical"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL80979.1"
/db_xref="GI:18892888"
/translation="MVAVRSFGEALDFLVKYLKLLLTIIIPALJSGVASPFVEEYEP
QRFDIIERYGAISPENKWLIPYVILVTVILSGIEFGIIDPAIKAKSGKDKFSL
KEIAIYGURNAGWGLLLTIISLLYLLALVIGGSIPTTILLLVDFIVDFLEPLVIVEL

```


AX339683 1134 bp DNA linear PAT 10-JAN-2002
LOCUS
DEFINITION Sequence 4 from Patent WO0196382.
AX339683
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pyrococcus furiosus
Pyrococcus furiosus
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
1
REFERENCE
AUTHORS Wicher, K.B., Holst, O.P., Hachem, M.Y., Karlsson, E.M. and
Hreggvidsson, G.O.
TITLE Thermostable cellulase
JOURNAL Patent: WO 0196382-A 4 20-DEC-2001;
Prokaria ehf. (IS)
FEATURES Location/Qualifiers
1..1134
/organism="Pyrococcus furiosus"
/mol_type="unassigned DNA"
/db_xref="taxon:2361"
82..1041
source
exon
ORIGIN

Query Match	99.8%;	Score	958.4;	DB	6;	Length	1134;
Best Local Similarity	99.9%;	Pred. No.	1e-254;				
Matches	959;	Conservative	0;	Mismatches	1;	Indels	0;
Qy	1	ATGAGCAAGAAAAGTTGCTCGTATCTATCTTAACAATCTCTTTTAGTACAGGCAATA	60				
Db	82	ATGAGCAAGAAAAGTTGCTCGTATCTATCTTAACAATCTCTTTTAGTACAGGCAATA	141				
Qy	61	TATTTTGTAGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA	120				
Db	142	TATTTTGTAGAAAAGTATCATCTCTGAGGCAAGTCAACTTCAATACCTCATCTACA	201				
Qy	121	CCACCCAAACAACACTTTTCCACTACCAAGGTTCTCAAGATTAGATACCTGTAGACGGT	180				
Db	202	CCACCCAAACAACACTTTTCCACTACCAAGGTTCTCAAGATTAGATACCTGTAGACGGT	261				
Qy	181	GAGTGGCCAGGAGCTCTATTGTATAAGGATGGTGATGGGAACCCAGAAATTTCTACATTGAA	240				
Db	262	GAGTGGCCAGGAGCTCTATTGTATAAGGATGGTGATGGGAACCCAGAAATTTCTACATTGAA	321				
Qy	241	ATAAACCTATGGAAACATCTTAATGCTACTGGATTTCCTGAGATGAGCTACAATTTAAACC	300				
Db	322	ATAAACCTATGGAAACATCTTAATGCTACTGGATTTCCTGAGATGAGCTACAATTTAAACC	381				
Qy	301	AGCGGCTCTCTCACTACGTCCTCAACACTTTGACAACTTGTCTTTGAGGATAGAAGTAAT	360				
Db	382	AGCGGCTCTCTCACTACGTCCTCAACACTTTGACAACTTGTCTTTGAGGATAGAAGTAAT	441				
Qy	361	TGGGTGATGGATACCCCGAAATATTCTATGGAAAACAGCCATGGGAATGCAAACTACGCA	420				
Db	442	TGGGTGATGGATACCCCGAAATATTCTATGGAAAACAGCCATGGGAATGCAAACTACGCA	501				
Qy	421	ACTGATGGCCCAATACCATTAACCCAGTAAAGTTTCAAACTTAACAGACTTCTATCTAACA	480				
Db	502	ACTGATGGCCCAATACCATTAACCCAGTAAAGTTTCAAACTTAACAGACTTCTATCTAACA	561				
Qy	481	ATCTCTTATAAACTTTGAGCCCAAGAACCGGCTGCGCAATTAACCTTCGCAATAGAAATCTCG	540				
Db	562	ATCTCTTATAAACTTTGAGCCCAAGAACCGGCTGCGCAATTAACCTTCGCAATAGAAATCTCG	621				
Qy	541	TTAACGAGAGAGCTTTGGAGAACACAGGGAATTAACAGCGATGAGCAGGAAGTAATGATA	600				
Db	622	TTAACGAGAGAGCTTTGGAGAACACAGGGAATTAACAGCGATGAGCAGGAAGTAATGATA	681				
Qy	601	TGGATTACTATGACGGATTACAAACCGCTGGCTCCAAAGTTAAGGAGATTGTAGTCCCA	660				
Db	682	TGGATTACTATGACGGATTACAAACCGCTGGCTCCAAAGTTAAGGAGATTGTAGTCCCA	741				
Qy	661	ATAATAGTTAAACGAAACACAGTAAATGTCTACATTTGAAATGTGGAAGGCAACATTGGT	720				
Db	742	ATAATAGTTAAACGAAACACAGTAAATGTCTACATTTGAAATGTGGAAGGCAACATTGGT	801				
Qy	721	TGGGAGTATGTTGCATTTAGAAATAAGACACCCCAATCAAGAGGGAAACAGTGAACAATTCGA	780				
Db	802	TGGGAGTATGTTGCATTTAGAAATAAGACACCCCAATCAAGAGGGAAACAGTGAACAATTCGA	861				
Qy	781	TACGGACATTTATPAGTGTTCGAGCCAAACATTTCAAGCTTACCAAATTAACAGAACTT	840				
Db	862	TACGGACATTTATPAGTGTTCGAGCCAAACATTTCAAGCTTACCAAATTAACAGAACTT	921				
Qy	841	TACTTAGAGGACGTGGAGATTGGAACTGAGTTTGGAAACCGCAACACTACTCTCGCCCAAC	900				
Db	922	TACTTAGAGGACGTGGAGATTGGAACTGAGTTTGGAAACCGCAACACTACTCTCGCCCAAC	981				
Qy	901	CTAGAGTGTGGATACAAAACATAAACACTAACTCCTCTAGATAGACCTTTATTTCCTAA	960				
Db	982	CTAGAGTGTGGATACAAAACATAAACACTAACTCCTCTAGATAGACCTTTATTTCCTAA	1041				
RESULT 7							
BD023645							
LOCUS	BD023645	867 bp	DNA	linear	PAT	27-AUG-2002	

RESULT 7
BD023645
LOCUS

QY 770 TGACAAATTCATACGAGCATTTATA 796
 Db 674 TAAATAACCAATAAATGAATTTGA 700

RESULT 8
 TN093354
 LOCUS
 DEFINITION
 Thermotoga neapolitana endo-1,4-beta-glucanase A (celA) and endo-1,4-beta-glucanase B (celB) genes, complete cds.
 U93354
 VERSION
 U93354.1 GI:1934672
 SOURCE
 Thermotoga neapolitana
 ORGANISM
 Thermotoga neapolitana
 Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 REFERENCE
 1 (bases 1 to 2803)
 BOK, J.D., YERNOL, D.A. and EVELEIGH, D.E.
 Purification, characterization, and molecular analysis of thermostable cellulases CelA and CelB from Thermotoga neapolitana Appl. Environ. Microbiol. 64 (12), 4774-4781 (1998)
 JOURNAL
 MEDLINE
 99054912
 PUBMED
 9835561
 REFERENCE
 2 (bases 1 to 2803)
 BOK, J.D. and EVELEIGH, D.E.
 Direct Submission
 TITL
 Submitted (13-MAR-1997) Biochemistry and Microbiology, Cook College, Rutgers University, Lipman Drive, New Brunswick, NJ 08903, USA

FEATURES
 source
 1..2803
 /organism="Thermotoga neapolitana"
 /mol_type="genomic DNA"
 /db_xref="taxon:2337"
 1..774
 /gene="celA"
 /EC_number="3.2.1.4"
 /function="hydrolysis of carboxymethyl cellulose, beta glucan"
 /note="TNCEL A"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="endo-1,4-beta-glucanase A"
 /protein_id="AAC95059.1"
 /db_xref="GI:1934673"
 /translation="MVELTAPGTADFRNDMPISMELNLNVRVYGTVMVRPDGERLTFNGDVEDLAREPERYILGYKFERHAAAGTKLPLVSVSVSTVELSFEIDHMSPLFMAMETWIRKYOVEASIGDVEIMVYFNELTGPKGKGYEYVSEFELNGEHRGIWELHAEKNNDYLAPRLKNPVRKRVFNVDKFLDVAGEYLSRSTRVKDFDLVFTVWEIGTERGSPETKSARFGWTFNFSIDMEVKG"
 759..1595
 /gene="celB"
 759..765
 /gene="celB"
 771..1595
 /gene="celB"
 /EC_number="3.2.1.4"
 /function="hydrolysis of carboxymethyl cellulose, beta glucan"
 /note="TNCEL B"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="endo-1,4-beta-glucanase B"
 /protein_id="AAC95060.1"
 /db_xref="GI:1934674"
 /translation="MRVLVSVFLVWVSAFLPSAEVWLTDIGATDITKGPVTVNELNFWNVKSYEGETWLFKDGKQVYADINVLQNPDSVHVGYPEIYGYKPPWAAHNSGTETLPVKVQLPDFVYVTLDSIWIENDLPINLAMETWITRKPDQTSVSSGDVEIMVWFYNN

ILMPGQKVDEFTTTIBINGSVPVETKWDVYFAPWGDVYLAFLRLTTPMKDGRVKNVD
 FVEKAAEVIKHKSTREVENFDEMYFCWSEIGTEFGDENTTAARFGWTFKDFSEIGE"

ORIGIN
 Query Match 9.3%; Score 89.2; DB 1; Length 2803;
 Best Local Similarity 49.8%; Pred. No. 1.5e-13;
 Matches 325; Conservative 0; Mismatches 303; Indels 24; Gaps 3;

QY 280 GAGATGACGTACAATTAACGAGCGGCTTCCTACTACGTCCACAACTTGACAAATTT 339
 Db 924 GAAAGCTGGCTCAAGTTCGATGGTGAAGAGGTCCAGTTCCTACGCGGATATCTACAACATC 983
 QY 340 GTCTTGAGGGATAGAAGTAATTTGGTGCATGATACCCGGAATATTTATGGAACAAG 399
 Db 984 GTCTTCAAAACCCGACAGCTGGGTTACGGCTATCCAGAGATATATACGGTACAG 1043
 QY 400 CCATGGAATGCAAACTACGCAACTGATGCCCAATACCATTTACCCAGTAAAGTTTCAAAAC 459
 Db 1044 CCGTGGCGGCGCACAAC-----AGCGGAACAGAGATTTCTCCGGTGAAGTGAAGAT 1097
 QY 460 CTAACAGACTTCTATCTAACATCTCTTAACTTGAGCCCAAGAGCGGCTGCCAATT 519
 Db 1098 CTTCTGACTTTTACGTAAACCTTGATTTCAATCTGTTACGAGAACGACCTTCCGATC 1157
 QY 520 AACTTCGCAATAGAATCTCTGTTAAACGAGAGAGCTTGGAGAACCAACAGGAATTAACAGC 579
 Db 1158 AACCTTGCATGGAACGTTGGATCAACAAGAACGACATCAGATCTG---TTTCTCTCT 1214
 QY 580 GATGACGAAGAGTAATGATATGATGATTTACTATGAGGATTAACACCGGCTGGCTCCAAA 639
 Db 1215 GGAGACGTGAGATCATGCTCTGTTCTTACAACCAATCTCTGATGCCAGGTGGCAGAG 1274
 QY 640 GTTAAAGGAGATGTAGTCCCAATATAGTTAAACGACACACAGTAAATGCTACATTTGAA 699
 Db 1275 GTGGATGATTTACAACACACGATCGAGATAAAGCGCTCTCTGTGGAACAATAATGGAC 1334
 QY 700 GTATGGAAGGCAACATTTGGTTGGAGTATGTTGATTTAGATTAAGAACCCCAATCAA 759
 Db 1335 GTTTACTTTTGGCCCTGGGGGCTGGGATTTATCTCGCTTCAGGCTCACAACTCCCATGAA 1394
 QY 760 GAGGACACAGTGACAAATTTCCATACGAGCATTTTAAAGTTTGGAGCCCAACATTT--- 814
 Db 1395 GATGCAGGGTGAAGTTCAACGTGAAGGACTTTTGGAAAAGCAGCGGAAGTTATCAA 1454
 QY 815 -----CAAGCTTACCAAAATTACAGAACTTTTACTTAGAGAGCTGGAGATTGGA 864
 Db 1455 AAACACTCAACAAAGGGTGGAAAACCTTCGATGAGATGATCTTCTCGCTCTGGGAGATAGGA 1514
 QY 865 ACTGAGTTTGGACCGCAGACACTACCTCCGCCACCTAGAGTGGTGGATCA 916
 Db 1515 ACAGAGTTCCGTTGATCCAAATACCAACCGCGCAAAATTTGGATGAGCACTTCA 1566

RESULT 9
 TNCEL A
 LOCUS
 DEFINITION
 T.maritima celA and celB gene.
 ACCESSION
 Z69341
 VERSION
 Z69341.1 GI:1297060
 KEYWORDS
 celA gene; celB gene; cellulase; endo-1,4-beta-glucanase.
 SOURCE
 Thermotoga maritima
 ORGANISM
 Thermotoga maritima
 Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 REFERENCE
 1 (bases 1 to 2800)
 LIEBL, W., RUILE, P., BRONNEMEIER, K., RIEDEL, K., LOTTSPESCH, F. and GREIF, I.
 TITL
 Analysis of a Thermotoga maritima DNA fragment encoding two similar thermostable cellulases, CelA and CelB, and characterization of the recombinant enzymes
 JOURNAL
 Microbiology 142, 2532-2542 (1996)
 REFERENCE
 2 (bases 1 to 2800)
 LIEBL, W.
 TITL
 Direct Submission

JOURNAL Submitted (02-FEB-1996) Liebl W., Technische Universitaet Muenchen, Lehrstuhl f. Mikrobiologie, Arcisstr. 21, 80290 Muenchen, Federal Republic of Germany

FEATURES Location/Qualifiers

source 1..2800
 /organism="Thermotoga maritima"
 /mol_type="genomic DNA"
 /strain="MSB8 (DSM3109)"
 /db_xref="taxon:2336"
 406..968
 repeat_region

/citation=[1]
 /rpt_type=OTHER
 /rpt_unit="406..435"
 1091..1101
 /citation=[1]
 1107..1880
 /gene="celA"
 1107..1880
 /gene="celA"

/EC number="3.2.1.4"
 /function="hydrolysis of carboxymethyl cellulose, beta-glucan, etc."
 /standard_name="cellulase"
 /citation=[1]
 /codon_start=1
 /evidence=experimental
 /transl_table=1
 /product="endo-1,4-beta-glucanase"
 /protein_id="CAA93273.1"
 /db_xref="GI:1297061"
 /db_xref="GOA:Q60032"
 /db_xref="SPTRMBL:Q60032"
 /translation="MVLMTKPGTSDFTVWNGIPLSMELNLWNIKEYSGSVAMKPDGKI
 TPADONLSPKPERVLCYPRFYGYKPEWENHTAEGSKLPVPSMKSEFSEVVSFD
 IHEPSLPLNFAETWLTREKYQTEASIGDVEIMVWFYNNLTTPGGEKLEETIPVL
 NGSEVGTWELAEWGDYLAFLKDFPVKGRKEDVFRHFLDAAGLAUSSARVKDF
 EDLYFTWEIGTEFGSPETKSAQFGWKFNFSIDLEVR"

1862..1872
 /gene="celA"
 /citation=[1]
 1877..2701
 /gene="celB"
 1877..2701
 /gene="celB"

/EC number="3.2.1.4"
 /function="hydrolysis of carboxymethyl cellulose, beta-glucan, etc."
 /standard_name="cellulase"
 /citation=[1]
 /codon_start=1
 /evidence=experimental
 /transl_table=1
 /product="endo-1,4-beta-glucanase"
 /protein_id="CAA93274.1"
 /db_xref="GI:1297062"
 /db_xref="GOA:Q60033"
 /db_xref="SPTRMBL:Q60033"
 /translation="MRWAVLLMVVFSALLFSRSVLTSGATDISFNGFPVTVMLNFW
 NVKSYGETWLDGEKVEFYADYLVNQPSWVHGYPEIYGYKWPAGHNSGVF
 LVPVKDLPFYVTVLDYSIWVNNLPINLAMETWITRSPDQTSVSGDAEIMVWFYNN
 VLMGGOKDFEFTTVINGVKQETWDVYFAPKWDYLAFLRTTPMKEGKVKINVD
 FVQAAVVKKHSTRINDFELFCWKEITGEPDNTTAAKGTWFRFSPVEVWK"

1877..1927
 /gene="celB"
 /citation=[1]
 1928..2698
 /gene="celB"
 /product="endo-1,4-beta-glucanase"
 2722..2744
 /citation=[1]

terminator

sig_peptide

mat_peptide

terminator

terminator

terminator

terminator

terminator

terminator

terminator

terminator

terminator

Best Local Similarity 50.2%; Pred. No. 3.6e-12;
 Matches 305; Conservative 0; Mismatches 278; Indels 24; Gaps 3;

QY 332 ACAACATTGCTCTTGAGGATAGAAAGTAATTGGGTGCATGGATACCCCGAAATATTCTATG 391
 |||||
 Db 2082 ACAACATCGTTCTTCAGAAATCCAGACAGCTGGTGCATGGATATCCAGATCTACTACG 2141

QY 392 GAAACAAGCATGGAATGCAAACTACGCAACTGATGGCCCAATACCATTCACCACTAAG 451
 |||||
 Db 2142 GTTACAAGCCCTGGCGGGGCACACAGCGGTGTTGAATTTCTCTCTG-----AAGG 2195

QY 452 TTTCAAACTAACAGACTTCTATCTAAACAATCTCTATAAACTTCGAGCCCAAGACGGCC 511
 |||||
 Db 2196 TGAAGATCTTCGGGATTCTACGTGACTCTTGATTTACTCGATCTGGTACGAACAAC 2255

QY 512 TCCCAATTAACTTCGCAATAGAAATCCTGGTTAAACAGAGAGACTTGGAGAACACAGGAA 571
 |||||
 Db 2256 TGCCTATCAACCTTGCCTATGGAGAGCTGGATCAGCAGAAAGTCCCGACAGACCTCTG--- 2312

QY 572 TTAACAGGATGAGCAAGAAGTAATGATATGATTTACTATGACGATTACAAACCGGCTG 631
 |||||
 Db 2313 TTTCTTCGGTGACCGGAGATCATGTTGGTCTCAACAACGTTCTGATGCCGCG 2372

QY 632 GCTCCAAAGTTAAGGAGATTGTAGTCCCAATAATAGTTAACGGAAACACAGTAAATGTA 691
 |||||
 Db 2373 GTCAGAAAGTGGACGAGTTCACACACAGTCGAGATAAAGCGCGTGAAGCAGGAAACAA 2432

QY 692 CATTGTAAGTATGGAAGGCAACATTTGTTGGGATGTTGTCATTTAGATAAAGACCC 751
 |||||
 Db 2433 AGTGGGATGTTTACTTTGACCGTGGGATGGGATTTACTCTTCTTACAGCTGACAAAC 2492

QY 752 CAATCAAGAGAGGAAACAGTGCATTTCCATACGAGAGCTTTTATAAGTGTTCGAGCAACA 811
 |||||
 Db 2493 CGATGAAGAAGAAAGGTGAAAAACAAGTGAAGGACTTCGTTTCAGAAAAGCGCGGAAG 2552

QY 812 TT-----TCAAGCTTACCAATTACAGAACTTTTACTTAGAGGCGGCG 856
 |||||
 Db 2553 TTGTCAAGAGCACTCAACAGAGATAGACAAATTCGAAGAGCTGATTTCTCGCTGGG 2612

QY 857 AGATTGGAACCTGAGTTTGGAAACGCCAAGCACTACCTCCGCCCCACCTAGAGTGGTGATCA 916
 |||||
 Db 2613 AGATCGGACGGAGTTTGGAGATCCGACACACAGCGCGGAAATTCGCTGGACTTCA 2672

QY 917 CAAACAT 923
 |||||
 Db 2673 GGGACTT 2679

RESULT 10
 AE001800 20653 bp DNA linear BCT 02-JUN-1999

LOCUS Thermotoga maritima section 112 of 136 of the complete genome.
 DEFINITION
 AE001800 AE000512
 ACCESSION
 AE001800.1 GI:4982090

VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Thermotoga maritima
 Thermotoga maritima
 Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.

REFERENCE 1 (bases 1 to 20653)
 Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
 Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
 McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
 Stewart,A.N., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
 Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
 Eisen,J.A., Fraser,C.M. et al.

Evidence for lateral gene transfer between Archaea and bacteria
 from genome sequence of Thermotoga maritima
 Nature 399 (6734), 323-329 (1999)

JOURNAL 98287316
 MEDLINE 10360571
 PUBMED 2 (bases 1 to 20653)
 REFERENCE 2 (bases 1 to 20653)
 Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
 Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,

McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Corton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Reidelberg, J., Sutton, G.G., Fleischmann, R.D., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES

source

1. 20853
/organism="Thermotoga maritima"
/mol_type="genomic DNA"
/db_xref="taxon:2336"
840. .1616
/gene="TM1524"
840. .1616
/gene="TM1524"
/note="similar to PID:1297061 GB:AE000512 percent identity: 100.00; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase"
/protein_id="AAD36591.1"
/db_xref="GI:4982091"

/translation="MVMIMTKPGTSDFWNGIPLSMELNLWNKEYSGSVAMKFDGEK
ITFDADIQNSPKPEERYVLGYPFYGYKPENHTAEGSKLPVPSSMKSPSVSEVF
DIHPELPLNFAMETLREKYQTEASIGDVEIMVMFYFNLTGGEKIEETIPFV
LNGESVGTWLAEBWGDYLAFLKDPVKGRKVDVRFHFLDAAGKALLSSARVKD
FEDLYFTWEGTGFSGPETSQAQFGKNFNSIDLEVR"
1613. 2437
/gene="TM1525"
1613. 2437
/gene="TM1525"
/note="similar to GP:1297062 percent identity: 100.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase"
/protein_id="AAD36592.1"
/db_xref="GI:4982092"

/translation="MRWALIMVFSALLFSSEVYLTSGATDISNPGFPVTWELNFW
NVKSYEGETWMLKDGKVEFYADLVNLPQDSWVHGYPEIYGYKPMAGHNSGVFE
LPVKXDLPPDYVLTLYSIWYNNLPLINLAETWITRSPDTSVSSGDAIMVMFYNN
VLMGQGVQDEFTTVEINGVKQTKWDVYFAPGWDYLAFLITPMKEGKVKINVD
FVQAAKAVKXSRIDNFBELIFCVWEIGTEGFDNTTAAKFGWTFRDSVEVK"
2557. 2994
/gene="TM1526"
2557. 2994
/gene="TM1526"
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36593.1"
/db_xref="GI:4982093"

/translation="MKVSDILTVAILREERGFYRLSHFNGEIKTKTLELADOER
IHAFIRKMSDOENWSDVSLAGYAFVEFDPDTEILRSKDLTKVELVDIAISVEKD
SIIILYELKGLVNSDAQTKVKLIIDQEKHLKLEMKKEST"
complement(2956. .33778)
/gene="TM1527"
complement(2956. .33778)
/gene="TM1527"
/note="similar to GB:AE000657 percent identity: 54.92;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAD36594.1"
/db_xref="GI:4982094"

/translation="MPGCKOKHVRALLRLVLSLENRPMSSSELAKELITRNPAK
VLHYLRKSLGWSIKGNKGRFLAKSPDEISFLDVISVEKDLVIVDCNSRCNKYRAC
RVKYFNWLSYDLKDFSNITKDIASGTFTHLEKLS"

gene

CDS

3406. .4275
/gene="TM1528"
3406. .4275
/gene="TM1528"
/note="similar to GB:X73124 SP:P39582 PID:413930
GB:AJ009126 percent identity: 57.75; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="1,4-dihydroxy-2-naphthoate
octaprenyltransferase, putative"
/protein_id="AAD36595.1"
/db_xref="GI:4982095"

/translation="MKYFVAVRPFVSASFVTVGAMLAEEFLIYILSIASIFL
HAGVNTNDYDKYKGVTDKDSGLSSGLLVSKITPRELLSVFCYAVSVLGLVLI
KISGALLWGLTGLVGYAVTGHPFYLYKSLGMLFVILMGPMLVLYAVVOTGRF
SLEALLSVIPYGIATDILLANEIRDSEFDRSGIKTLPILIGDRAASFVAVTGLI
YVFVILVSTGVFRVLSLVSLVPLTIRVIKQLFQKSAGRNAREIADVDKMSALAE
MLFVSMILGLLR"
4275. .4901
/gene="TM1529"
4275. .4901
/gene="TM1529"
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36596.1"
/db_xref="GI:4982096"

/translation="MPVVKRISSEISFLSLGLFLGMVYFQVNMDFWMEFMTTLL
RSVISIFLGGKRNKPNKYEARPLAYSNAAAYLLFVYGLISLLIPFHRDVSVYGL
AEGNALKVILLVFAFFETIWRGFVFEFLQRLDVPVPSILSSIVSVHFTGN
NALIVAFVILGIIISLLYITVKVSTPAFAHALWSLIFVVLPLKGGF"
5003. .5860
/gene="TM1530"
5003. .5860
/gene="TM1530"
/note="similar to PID:1055220 SP:P52040 percent identity:
66.53; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="electron transfer flavoprotein, beta subunit"
/protein_id="AAD36597.1"
/db_xref="GI:4982097"

/translation="MNVVVCIKQVPTNVRIDRKNLVREGVPSIINPDRALEL
ASQKKEFGATVYITMGPPQAKKADATAFGLDEAVHLSDRTPAGADTLATYTL
WGIKIEIRIKIDILITLTKQAVDGTQGVGLATRFYALGAYVVRIBEDPEKKE
NVIVRLDQGFELRLKLPVLTITDELNRPYADLPNLIRAIRYEPVIVTHKDLGLD
PKKCGFGSPTRVVTNIPPAKGGDIISKNEDEPEVAEKLEALKKKFEAVRLVEALK
FVLEGEKDE"
5853. .6869
/gene="TM1531"
5853. .6869
/gene="TM1531"
/note="similar to PID:1055221 SP:P52039 percent identity:
68.15; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="electron transfer flavoprotein, alpha subunit"
/protein_id="AAD36598.1"
/db_xref="GI:4982098"

/translation="MSEKKIIFVLEIHHGKHAHPVSWELIGKARDLASKLNSVWGV
LIGELGESVAEKAIQRGADKVLVYKRNRENTYVNYLYKALVDVRYRPEIFLIGAT
LEGRLAGWATELETADTGLDIIIPDKKLLAMTRPTFGNLMATIMCPDRPQM
ATVRPGVMKELPPDPTGTEIIEEYDITGTFDKLIEILETIPLOTQNVLEVPVAVQ
CKGVGPGEPFKLLEADLLGGEYGASAAVAGWISPEHOVSQGTGTVRVLVYACG
ISGAQHVVGINESIIVAINIDEKAPFDIADIGVDLHKVVPALTAKLRELINKS
GVKK"
6866. .8182
/gene="TM1532"
6866. .8182
/gene="TM1532"

gene

CDS

3406. .4275
/gene="TM1528"
3406. .4275
/gene="TM1528"
/note="similar to GB:X73124 SP:P39582 PID:413930
GB:AJ009126 percent identity: 57.75; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="1,4-dihydroxy-2-naphthoate
octaprenyltransferase, putative"
/protein_id="AAD36595.1"
/db_xref="GI:4982095"

/translation="MKYFVAVRPFVSASFVTVGAMLAEEFLIYILSIASIFL
HAGVNTNDYDKYKGVTDKDSGLSSGLLVSKITPRELLSVFCYAVSVLGLVLI
KISGALLWGLTGLVGYAVTGHPFYLYKSLGMLFVILMGPMLVLYAVVOTGRF
SLEALLSVIPYGIATDILLANEIRDSEFDRSGIKTLPILIGDRAASFVAVTGLI
YVFVILVSTGVFRVLSLVSLVPLTIRVIKQLFQKSAGRNAREIADVDKMSALAE
MLFVSMILGLLR"
4275. .4901
/gene="TM1529"
4275. .4901
/gene="TM1529"
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36596.1"
/db_xref="GI:4982096"

/translation="MPVVKRISSEISFLSLGLFLGMVYFQVNMDFWMEFMTTLL
RSVISIFLGGKRNKPNKYEARPLAYSNAAAYLLFVYGLISLLIPFHRDVSVYGL
AEGNALKVILLVFAFFETIWRGFVFEFLQRLDVPVPSILSSIVSVHFTGN
NALIVAFVILGIIISLLYITVKVSTPAFAHALWSLIFVVLPLKGGF"
5003. .5860
/gene="TM1530"
5003. .5860
/gene="TM1530"
/note="similar to PID:1055220 SP:P52040 percent identity:
66.53; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="electron transfer flavoprotein, beta subunit"
/protein_id="AAD36597.1"
/db_xref="GI:4982097"

/translation="MNVVVCIKQVPTNVRIDRKNLVREGVPSIINPDRALEL
ASQKKEFGATVYITMGPPQAKKADATAFGLDEAVHLSDRTPAGADTLATYTL
WGIKIEIRIKIDILITLTKQAVDGTQGVGLATRFYALGAYVVRIBEDPEKKE
NVIVRLDQGFELRLKLPVLTITDELNRPYADLPNLIRAIRYEPVIVTHKDLGLD
PKKCGFGSPTRVVTNIPPAKGGDIISKNEDEPEVAEKLEALKKKFEAVRLVEALK
FVLEGEKDE"
5853. .6869
/gene="TM1531"
5853. .6869
/gene="TM1531"
/note="similar to PID:1055221 SP:P52039 percent identity:
68.15; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="electron transfer flavoprotein, alpha subunit"
/protein_id="AAD36598.1"
/db_xref="GI:4982098"

/translation="MSEKKIIFVLEIHHGKHAHPVSWELIGKARDLASKLNSVWGV
LIGELGESVAEKAIQRGADKVLVYKRNRENTYVNYLYKALVDVRYRPEIFLIGAT
LEGRLAGWATELETADTGLDIIIPDKKLLAMTRPTFGNLMATIMCPDRPQM
ATVRPGVMKELPPDPTGTEIIEEYDITGTFDKLIEILETIPLOTQNVLEVPVAVQ
CKGVGPGEPFKLLEADLLGGEYGASAAVAGWISPEHOVSQGTGTVRVLVYACG
ISGAQHVVGINESIIVAINIDEKAPFDIADIGVDLHKVVPALTAKLRELINKS
GVKK"
6866. .8182
/gene="TM1532"
6866. .8182
/gene="TM1532"

		/note="similar to SP:P26484 GB:X55450 PID:38698 percent identity: 66.82; identified by sequence similarity; putative"	
		/transl_table=1	
		/product="fixC protein"	
		/protein_id="AAD36599.1"	
		/db_xref="GI:4982099"	
		/translation="MKIEFDVVVVGAGPSGLSAYVLAKNGLKVAIVVEKGEVPGSKNP MGCVLHPHLEKTEPFLFEKAANSKALERNV;BQNLLWLLGNEGVILGHRNVSEKNP NATPFLRAMPDFRFAVEVEKAGALIIPKTVFEDFLRNEKEGIAGVTVSRPFGKEIHSKA VTAEGVNPFLTWAGLRKEDLKPWVAIVAVKEVISVPEDVNRVRFVSGNDGATLEL LGSNSGEMGMFLVNRSSVSLGCVLLEDLRKKIKPVQLLENLKNHPVITSMDLGE YRNTWEYLAHLIPGEGYAMPKPYGDRVLVCGDAAMLVNS;IHREGSNHAI;TSGRLAA ETULFEAGDFPEKILKNT;LLKKEFSILKLEKTKDLMPTMEKHKQFVFIYPDLAN DALKRLQVDGTPKVDQKQIADWLSRSLISIGISLDLLRFWEAVR"	
	gene	8189. .8467	
		/gene="TM1533"	
	CDS	8189. .8467	
		/gene="TM1533"	
		/note="similar to GB:X65515 PID:510488 SP:P53658 percent identity: 68.82; identified by sequence similarity; putative"	
		/codon_start=1	
		/transl_table=1	
		/product="ferredoxin"	
		/protein_id="AAD36600.1"	
		/db_xref="GI:4982100"	
		/translation="MKIEDKLYLNRYRTDEENPHLKIODESICAEKCSDEPCVSCCPA DVTWEYSGMEVFEGLCEGTCRIVCPPFNIEWNYPNGNVLVYKFG"	
	gene	8477. .8818	
		Query Match 8.8%; Score 84.2; DB 1; Length 20653;	
		Best Local Similarity 50.2%; Pred. No. 3,7e-12;	
		Matches 305; Conservative 0; Mismatches 278; Indels 24; Gaps 3;	
QY	332	ACAACATGTTGCTTGTAGGGATAGAAAGTAATTTGGGTGCATGGATACCCCGAAATATTTCTATG	391
Db	1818	ACAACATCGTTCTTCAGAAATCCAGACAGCTGGGTGCATGGATATCCAGAGATCTACTACG	1877
QY	392	GAACAGCGCATGGAATGCAAACTACGCCAACTGATGSCCAATAACCATTAACCCAGATAAG	451
Db	1878	GTTCACAGCCCTGGGGGGGCAACAGCGGTGTTGAAATTTCTTCCTGTG-----AAGG	1931
QY	452	TTTCAACACCTAACAGACTTCTATCTAAATCTCTATAAACTTGAGCCCAAGAACGGCC	511
Db	1932	TGAAAGATCTTCGGATTCTACGTGACTCTTGATTACTCGATCTGGTACGAAAACAAC	1991
QY	512	TGCCAATTAACCTTCGCAATAGATCTCTGGTTAACGAGAGAAGCTTGGAGAACACAGGAA	571
Db	1992	TGCTATCAACCTTGCCATCGAGACGTGGATACGAGAAGTCCGACCCAGACCTCTG---	2048
QY	572	TTAACACGCGATGACCAAGAAGTAATGATGGAATTTACTATGACGGATTACAAACCGCGTG	631
Db	2049	TTTCTTCGGGTGACCGGAGATCATGGTTGGTTCTACACAAAGTTCTGATGCCGGCG	2108
QY	632	GCTCCAAAGTTAAGGAGATTGTAGTCCCAATAATAGTTAAACGGAACACCAAGTAAATGCTA	691
Db	2109	GTCAGAAAGTGGACAGAGTTCACTACACAGCTCGAGATAAAGCGGCGTGAACACGAGAAACNA	2168
QY	692	CATTTTGAAGTATGGAAGGCAAAACATTCGTTGGGAGTATGTTGCAATTTAGATAAAGACCC	751
Db	2169	AGTGGGATGTTTACTTTTGACCGTGGGGATGGGATTAACCTTCTTCAGACTGCAACAC	2228
QY	752	CAATCAAGAAGGGGAACAGTGAACAATTCATACGGAGCATTTAATAGTGTGTGACGCCAAC	811
Db	2229	CGATGAAGAAGGAAGGTGAAATCAACGCTGAAGAGACTTCGTTTCAGAAAGCCGCGAAG	2288
QY	812	TT-----TCAAGCTTACCAATTAACAGACTTTTACTTAGAGGACCGTG	856
Db	2289	TTGTCAAGAGCACTCAACGAGAAATAGACAAATTTGGAAGAGCTGTATTTCTGCGCTCTGGG	2348
QY	857	AGATTGAACTGAGTTTGGAAACGCACAGCACTACTCTCGGCCCACTAGAGTGTGGATCA	916

Db 2349 AGATCGGACGGAGTTTGGAGATCCGAACACACAGCGGCAAAATTCGGCTGGACTTTCA 2400
 Qy 917 CAAACAT 923
 Db 2409 GGGACTT 2415

RESULT 11
 TNZ86103
 LOCUS T.neapolitana cela and celB genes and two open reading frames.
 DEFINITION
 ACCESSION Z86103
 VERSION Z86103.1 GI:1870178
 KEYWORDS cela gene; celB gene; cellulase; endo-1,4-beta-glucanase.
 SOURCE Thermotoga neapolitana
 ORGANISM Thermotoga neapolitana
 Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.

REFERENCE 1
 AUTHORS Dakhova,O.N., Kurepina,N.E., Zverlov,V.V., Svetlichnyi,V.A. and Velikodvorskaya,G.A.
 TITLE Cloning and expression in Escherichia coli of Thermotoga neapolitana genes coding for enzymes of carbohydrate substrate degradation
 JOURNAL Biochem. Biophys. Res. Commun. 194 (3), 1359-1364 (1993)
 MEDLINE 93356813
 PUBMED 8352795
 REMARK (sites)
 REFERENCE 2 (bases 1 to 4547)
 AUTHORS Zverlov,V.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 4547)
 AUTHORS Zverlov,V.
 TITLE Direct Submission
 JOURNAL Submitted (27-FEB-1997) Zverlov V., Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov Sq., Moscow, Russia, 123182

FEATURES
 source Location/Qualifiers
 1..4547
 /organism="Thermotoga neapolitana"
 /mol_type="genomic DNA"
 /strain="Z2706-MC24"
 /db_xref="taxon:2337"
 1..326
 /gene="orf1"
 /pseudo
 <1..326
 /gene="orf1"
 /pseudo
 /codon_start=3
 /transl_table=11
 /db_xref="PSEUDO:CAB06780.1"
 /db_xref="RENTREMBL:CAB06780"
 756..761
 767..2332
 /gene="orf2"
 767..2332
 /gene="orf2"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="CAB06781.1"
 /db_xref="GI:1870179"
 /db_xref="SPTREMBL:P96490"
 /translation="MKSIEGLKETFYKGFASLPANVMILLEENLPKECEVFLCILKLA
 WKGNFEAIQKDKLKSCKSEAKFLFLANKLAFKYTGKVDNNLYRYLKNLPKMS
 KSRDITVITFLINFEPTGICPKRKVRWKNDRKSTLSFLYLSLARREADSGRLSEAV
 HDYIQAYRLGVEPHPTCIYSSLNLDWDIRERKPKSLAHALSQAVFWLGYVEEPEEN
 LQALDITFLVVEKDMSPSHSTAHILVSLPVEDYLSLKKAKFVLDYTGSTPTN
 SOLRYYVEKVAWKGTSKSGISDILKGTGMIRADTIRKLLTSGVDTGAPFVWMEW
 IKWEIPKYESSEKLSLSEKTLPLSSETNPFPLTTYALLDREFLSRKELKRAYKLLEDI
 ESADFLWAKHRTWTFVSVNKAHPFIEGRKETVKKALGEMKPKRLERPTLRVLEWKE
 SDRKLDNFURNYSVREFRVGLKGPVVRRFARRISLKVQFLVAFWCEDEGRV

gene
 CDS
 RBS
 gene
 CDS

```
RRLETLRYMVQV"
3354..3358
3367..4140
/gene="celA"
3367..4140
/gene="celA"
/EC number="3.2.1.4"
/function="hydrolysis of beta-1,4-glucoside bonds"
/standard_name="cellulase"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="endo-1,4-beta-glucanase"
/protein_id="CAB06783.1"
/db_xref="GI:1870180"
/db_xref="GOA:P96491"
/db_xref="SPTREMBL:P96491"
/translation="MVELTAPGTADFWRNDPLSMELNWNVERYTGTVVVRFDGERL
TFNGDVEDLSAREPERVILGYPEFYGYKPERHAAGTKLPLVSVSEFTVELSEFL
IDHPSLPLNFATWLTREKYQVEASIGDVEIMWVFYFNEITPGGKRGVEYTVSTEL
NGEQRGTIEWHAEWMDYLAFRLKNPVRKRVRFNVKDFLDVAGEYLSRSTRVKDF
DLYFTVWEIGTERGSPETKSARFGWTFNNFSIDMEVKG"
4125..4129
/gene="celA"
4137..4547
/gene="celB"
4137..4547
/gene="celB"
/EC number="3.2.1.4"
/function="hydrolysis of beta-1,4-glucoside bonds"
/standard_name="cellulase"
/codon_start=1
/transl_table=11
/product="endo-1,4-beta-glucanase"
/protein_id="CAB06782.1"
/db_xref="GI:1870181"
/db_xref="GOA:P96492"
/db_xref="SPTREMBL:P96492"
/translation="MRLIVSFLVLSAFPLSAEVLVDIGATDITFKGPFVWMLNFW
NVKSYGETWIKPQGVQFYADYINVLQNPDSVHGYPEIYGYKPMWAHNSGTETI
LPVKVLPDFYVTLDSYIWNELPILAMETWI"

ORIGIN
Query Match 8.5%; Score 82; DB 1; Length 4547;
Best Local Similarity 49.2%; Pred. No. 1.5e-11;
Matches 350; Conservative 0; Mismatches 335; Indels 27; Gaps 4;

QY 230 TCTACATTGAATAAACCTATGGAACATCTTAAATGCTACTGCTGATGTTGCTGAGATGACGT 289
DB 3422 TTTCATGGAGCTGAATCTGTGGAACGTAGAAAGATACACGGGAACGGTGGTGATGAT 3481
QY 290 ACAATTTAACAGGGGCTCTTCACTACGTCCCAACACTTGACAACTTGCTGAGGG 349
DB 3482 TCGA-----CGGTGAAGACTCACTTCAACGGTGATGTGTAAGATCTTTCTGCCAGAG 3535
QY 350 ATAGAAGTAATGGGTGATGATACCCCGAAATATCTATGTAACAAAGCCATGGAATG 409
DB 3556 AGCCAGAAGGTACATCTTGCTATCTGAGTTTACTACGGATACAAACCTGGGAA 3595
QY 410 CAACACTGCAACTGATGCGCCCAATACCATTAAGTTTAAACCTACAGACT 459
DB 3596 GACATGACGACGAGG-----GGACAAAGTTTACCTCTGCTAGTCTCTCTGCGAGTCCT 3649
QY 470 TCTATCTAACCAATCTCTATATACTTGAAGCCCAAGAACCGGCTCCCAATTAACCTGCCAA 529
DB 3650 TCACGGTCAACTTCTTCCGAATAGACCATGCGCTCTCTTCCGCTGACTTGCCA 3709
QY 530 TAGAATCTCTGTTAAACGAGAGAGCTTGGAGAAACACAGGAATTAACAGCGATGAGCAAG 589
DB 3710 TGGAAACCTGCTGCACAAAGAGAAAATACCAGGTCGAAGCTTCCATCGGAGATGTG---G 3766
QY 590 AAGTAATGATGATGATTTACTATACCGGATTAACACCGGCTGGCTCCCAAGTTTAAGAGA 649
```

```
DB 3767 AGATCATGCTGTGGTTCTATTCAACGAACCTCACACCGGTGGAAGAAAGTGGAGAGT 3826
QY 650 TTGTAGTCCCAATAATAGTTAACGGACACACAGTAAATGCTACATTTGAAGTATGGAAGG 709
DB 3827 ACACGGTTTGGTTCGAACCTTAACGGAGAACAAAAACGTGGAATCTGGGAACCTCTGGCAG 3886
QY 710 CAAACATTTGGTTGGGAGTATGTTGCATTTAGAAATAAAGACCCCAATCAAAAGAGGGAACAG 769
DB 3887 CCGAATGGAACCTGGGATTAATCTGTTCCGCTCAAGAAATCCTGTGAGAAAGGGAGG 3946
QY 770 TGACAATTCATACGGAGCATTTAAGTTGTGCGACCAACATTT-----CAA 817
DB 3947 TGAGATTCAACGTGAAGGATTTTCTCGATGTTGCGAGGAGATCTTTCAAGGTCCACAC 4006
QY 818 GCTTACCAAAATTACACAGAACTTTACTTAGAGAGCTGGAGATTTGGAACCTGAGTTTGGAA 877
DB 4007 GTGTGAAGGATTTTGATGATCTGTATTTACCGTCTGGAGATCGAACCGAATTTGGA 4066
QY 878 CGCCAAGCACTACTCCCGCCACCTAGAGTGTGATGATCAAAACATAACACT 929
DB 4067 GCCTGAAACAAAGAGTGCCTGTTTGGATGACGCTTCAATAACTTTTCAAT 4118

RESULT 12
166494/c
LOCUS 166494 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F., Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source
1..7218
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 5.6%; Score 53.8; DB 6; Length 7218;
Best Local Similarity 5.4%; Pred. No. 0.00099;
Matches 22; Conservative 218; Mismatches 165; Indels 0; Gaps 0;

QY 502 AAGAACGGCTCCCAATTAACCTTCGCAATAGAAATCCTGTTAAAGAGAGAGCTTGGAGA 561
DB 1447 AAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1388
QY 562 ACAACAGGAATTAACAGCGATGAGCAAGAGTAATGATATGATTTACTATGACGATTA 621
DB 1387 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1328
QY 622 CAACCGCTGGCTCCCAATTAAGAGAGATTTAGTCCCAATAATAGTTTAAACGCAACCA 681
DB 1327 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1268
QY 682 GTAATGCTACATTTGAAGTATGGAAGCAACATTTGTTGGAGTATGTTGCATTTAGA 741
DB 1267 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1208
QY 742 ATAAAGACCCCAATCAAGAGGAAACAGTCAATTCATACGAGACATTTAATAGTTT 801
DB 1207 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1148
QY 802 GCAGCAACATTTCAAGCTTACCAATTAACAGAACTTTTACTTAGAGAGCTGAGATTT 861
DB 1147 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1088
QY 862 GGAACCTGAGTTTGGAAACCCCAAGCACTACCTCCGCCACCTAGAG 906
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 20:10:44 ; Search time 22 seconds
(without alignments)
748.577 Million cell updates/sec

Title: US-09-914-543-46
Perfect score: 1701
Sequence: 1 MSKKFVIVSILITILLVQAI.....HLEWITNITLPLDRPLIS 319

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCUTS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	100.0	319	3	US-09-215-042-1
2	1701	100.0	319	4	US-09-134-078-64
3	486.5	28.6	288	3	US-08-985-280C-2
4	486.5	28.6	288	3	US-09-215-042-2
5	143.5	8.4	261	3	US-09-110-359A-2
6	128	7.5	260	3	US-09-216-295-23
7	128	7.5	260	4	US-09-632-570-23
8	128	7.5	260	4	US-09-632-575-53
9	119.5	7.0	553	1	US-08-651-572-2
10	119.5	7.0	553	3	US-09-066-544-2
11	119.5	7.0	553	3	US-08-981-086-2
12	119.5	7.0	553	4	US-09-430-669-2
13	114.5	6.7	381	3	US-09-216-295-22
14	114.5	6.7	381	4	US-09-632-570-22
15	114.5	6.7	381	4	US-09-632-575-52
16	114	6.7	233	1	US-08-032-848C-12
17	112.5	6.6	264	4	US-09-632-570-24
18	112.5	6.6	264	4	US-09-632-575-54
19	111	6.5	239	3	US-08-849-751-2
20	111	6.5	239	4	US-08-478-816-2
21	111	6.5	667	1	US-08-471-033-7
22	111	6.5	667	2	US-08-471-044-7
23	111	6.5	667	2	US-08-463-483A-7
24	111	6.5	667	2	US-08-471-046A-7
25	111	6.5	667	2	US-08-470-566B-7
26	111	6.5	667	2	US-08-469-334-7
27	111	6.5	667	3	US-09-300-529-7

28	108	6.3	263	3	US-09-216-295-24	Sequence 24, Appl
29	108	6.3	312	3	US-09-216-295-21	Sequence 21, Appl
30	108	6.3	371	3	US-09-104-308-1	Sequence 1, Appl
31	108	6.3	371	3	US-09-321-981-1	Sequence 1, Appl
32	108	6.3	371	4	US-09-739-861A-1	Sequence 1, Appl
33	108	6.3	371	4	US-09-795-583-1	Sequence 21, Appl
34	108	6.3	371	4	US-09-632-570-21	Sequence 21, Appl
35	108	6.3	371	4	US-09-632-575-51	Sequence 51, Appl
36	108	6.3	386	3	US-09-321-981-5	Sequence 5, Appl
37	108	6.3	386	4	US-09-739-861A-5	Sequence 5, Appl
38	108	6.3	386	4	US-09-795-583-5	Sequence 5, Appl
39	103	6.1	852	1	US-08-471-033-36	Sequence 36, Appl
40	103	6.1	852	2	US-08-471-044-36	Sequence 36, Appl
41	103	6.1	852	2	US-08-463-483A-36	Sequence 36, Appl
42	103	6.1	852	2	US-08-471-046A-36	Sequence 36, Appl
43	103	6.1	852	2	US-08-470-566B-36	Sequence 36, Appl
44	103	6.1	852	2	US-08-469-334-36	Sequence 36, Appl
45	103	6.1	852	3	US-09-300-529-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-215-042-1
; Sequence 1, Application US/09215042
; Patent No. 6426698
; GENERAL INFORMATION:
; APPLICANT: Liu, Jiyin
; TITLE OF INVENTION: Continuous Biopolishing of
; TITLE OF INVENTION: Cellulose-Containing Fabrics With Thermophilic Cellulases
; FILE REFERENCE: 5464.200-US
; CURRENT APPLICATION NUMBER: US/09/215,042
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/068,274
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-215-042-1

Query Match	100.0%;	Score 1701;	DB 3;	Length 319;
Best Local Similarity	100.0%;	Pred. No. 3.7e-167;	Mismatches 0;	Indels 0; Gaps 0;
Matches 319;	Conservative 0;			
QY	1	MSKKFVIVSILITILLVQAIYFVKYHTS	EDKSTNTSSTPQTTLSTTKVLKIRYPDDG	60
Db	1	MSKKFVIVSILITILLVQAIYFVKYHTS	EDKSTNTSSTPQTTLSTTKVLKIRYPDDG	60
QY	61	EWFGAPIDKDGNGNPEFYIEINLWLNATG	PAETYNLTSGVLHYVQOLDNIVLRDSN	120
Db	61	EWFGAPIDKDGNGNPEFYIEINLWLNATG	PAETYNLTSGVLHYVQOLDNIVLRDSN	120
QY	121	WHGYPEIFYGNKPNANVATDGPILPSKVN	LDFTLTSYKLEPKNGLPINFAISW	180
Db	121	WHGYPEIFYGNKPNANVATDGPILPSKVN	LDFTLTSYKLEPKNGLPINFAISW	180
QY	181	LTRAWRTTGINSDEQVMWIIYDGLQFAGS	KVKEIIVVPIIVNGTVPVNATPEVWKANIG	240
Db	181	LTRAWRTTGINSDEQVMWIIYDGLQFAGS	KVKEIIVVPIIVNGTVPVNATPEVWKANIG	240
QY	241	WEYVAFRIKTPKEGTVIPIYCAFI	SVAAINISLNYLYLEDVEICTEGCTBTSSAH	300
Db	241	WEYVAFRIKTPKEGTVIPIYCAFI	SVAAINISLNYLYLEDVEICTEGCTBTSSAH	300
QY	301	LEWMIITNITLPLDRPLIS	319	
Db	301	LEWMIITNITLPLDRPLIS	319	

```
RESULT 2
US-09-134-078-64
; Sequence 64, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: Fast-Seq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 319 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-134-078-64

Query Match 100.0%; Score 1701; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 3.7e-167;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKKFVIVSLTLLVQAIFYVEKYHTSBDKSTNTSSTPPQTTSTTKVLKIRYPDDG 60
Db 1 MSKKFVIVSLTLLVQAIFYVEKYHTSBDKSTNTSSTPPQTTSTTKVLKIRYPDDG 60

Qy 61 EWPQAPDKDGDGPEFYIEINLWNLNATGFAEMTYNLTSGVLHYVQQLDNIVLDRSN 120
Db 61 EWPQAPDKDGDGPEFYIEINLWNLNATGFAEMTYNLTSGVLHYVQQLDNIVLDRSN 120

Qy 121 WWHGYPEIFYGNKPWNANYATDGPILPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180
Db 121 WWHGYPEIFYGNKPWNANYATDGPILPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180

Qy 181 LTREAWTTGNSDEQEWMIWYVDGLOPAGSKVKEIWPVPIVNGTPVNAFTFVKANIG 240
Db 181 LTREAWTTGNSDEQEWMIWYVDGLOPAGSKVKEIWPVPIVNGTPVNAFTFVKANIG 240

Qy 241 WEYVAFRIKPIKEGTVTIPYGAFISVAANISSLPNTYLYLEDEVEIGTFGPTTSAH 300
Db 241 WEYVAFRIKPIKEGTVTIPYGAFISVAANISSLPNTYLYLEDEVEIGTFGPTTSAH 300

Qy 301 LEWMITNITLTPDRPLIS 319
Db 301 LEWMITNITLTPDRPLIS 319

RESULT 3
US-08-995-280C-2
; Sequence 2, Application US/08995280C
; Patent No. 6043075
; GENERAL INFORMATION:
; APPLICANT: Bjornvad, Mads Eskelund
; APPLICANT: Schulein, Martin
; APPLICANT: No. 6043075revang, Iben Angelica
; TITLE OF INVENTION: A No. 6043075el Endoglucanase
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6043075o No. 6043075disk of No. 6043075th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,280C
; FILING DATE: 19-Dec-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol A.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4906.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-995-280C-2

Query Match 28.6%; Score 486.5; DB 3; Length 288;
Best Local Similarity 32.2%; Pred. No. 7.6e-42;
Matches 104; Conservative 58; Mismatches 106; Indels 55; Gaps 5;

Qy 4 KKFVIVSLTLLVQAIFYVEKYHTSBDKSTNTSSTPPQTTSTTKVLKI 54
Db 2 KXSLSLILILLITLSFSQTPKYKADAFILKAPSSGDVTTKNLPLT 47

Qy 55 RYDDGEPGAPDKDGDGNEFYIEINLWNLNATGFAEMTYNLTSGVLHYVQQLDNIV 114
Db 48 -----LELFWNIANYEGNTWMAFYKEEDTVEYYADIKNIV 83

Qy 115 LRDRNNWVHGPEIFYGNKPWNANYATDGPILPSKVSNTLDFYLTISYKLEPKNGLPIN 174
Db 84 LKDKNSWVHGPEIFYGYKVPWAGHNSIEKLALPKVSEFPDVLNFKYNIWYKSLPIN 143

Qy 175 FAIESWLTREAWTTGNSDEQEWMIWYVDGLOPAGSKVKEIWPVPIVNGTPVNAFTFV 234
Db 144 FAMEWTIKPEYQKT-VTSGDIEMWMLYANLSPAGKVGKIPILNNGNQDKIWEV 202

Qy 235 WKANIGWYVAFRIKPIKEGTVTIPYGAFIS-----VAANISSLPNTYLYLEDEVEI 287
Db 203 YLSPMSWDYVAYKSKENILQGVKIPINEFLKRLTILANPNPSRITPEKFDQMTVTWVEI 262

Qy 288 GTEGCTPSTTSAHLEWMITNITL 310
Db 263 GTEFGDPYTTEAKFGWTFNFDI 285
```

```
RESULT 4
US-09-215-042-2
; Sequence 2, Application US/09215042
; Patent No. 6126698
; GENERAL INFORMATION:
; APPLICANT: Liu, Jiyin
; APPLICANT: Condon, Brian
; TITLE OF INVENTION: Continuous Biopolishing of
; TITLE OF INVENTION: Cellulose-Containing Fabrics With Thermophilic Cellulases
; FILE REFERENCE: S464.200-US
; CURRENT APPLICATION NUMBER: US/09/215.042
; CURRENT FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/068,274
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Dictyoglomus sp.
US-09-215-042-2

Query Match      28.6%; Score 486.5; DB 3; Length 288;
Best Local Similarity 32.2%; Pred. No. 7.6e-42;
Matches 104; Conservative 58; Mismatches 106; Indels 55; Gaps 5;

QY 4 KKFVIVSILITLLVQAIV-----FVEKYHTSEDKSTNTSSPTTTLSTTKYLKI 54
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 KKSLLSLLILLTLTSLFSQTPPKYKDAFILKAPSSGDVTTKNPLT----- 47

QY 55 RYPDGGWPGAPIDKDGNGNPEFYIENLNILNATGFAEMTYNLTSGVLHYVQQLDNIV 114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 48 -----LELFWNIANYEGNTWMAFYKEEDTVEYVADIKNIV 83

QY 115 LDRSNVHGYPEIFYGNKPNANYATDGPILPSPKSVNLTDFVLTISYKLEPKNGLPIN 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 LKXNSVHGYPEYVYGYKPGWAGHNSIEKIALPKKUSEPFDVLFNKNYIWEKULPIN 143

QY 175 FAIESWLTREAWRTTGINSDEQEVMIWIYDGLQPGAGSKYKEIVVPIIVNGTPVNAFVEM 234
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 144 FAMEWTITKEPYQKT-VTSGDIEMWMLYANRLSPAGRKVGEVKIPIILNGNQDIWEV 202

QY 235 WKANIGHEVAFRIKTIKEGTVTIPYGATIS-----VAANISSL--PNYTELYLDEVEI 287
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 203 YLSPMSWDYVAYKSENILQGVKIPINEFLUKHLRTILANNPSCRITPEKFDQMYVTWEI 262

QY 288 GTEFGTPTTSAHLEWITNITL 310
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 263 GTEFGDPVTTAKGWTFSNFDI 285

RESULT 5
US-09-110-959A-2
; Sequence 2, Application US/09110959A
; Patent No. 6268197
; GENERAL INFORMATION:
; APPLICANT: Schulein, Martin
; APPLICANT: Outtrup, Helie
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Bjornvad, Mads Eskelund
; TITLE OF INVENTION: Alkaline Xyloglucanase
; FILE REFERENCE: 5206.200-US
; CURRENT APPLICATION NUMBER: US/09/110.959A
; CURRENT FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 0822/97
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: 1213/97
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/054,039
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 60/063,694
; PRIOR FILING DATE: 1997-10-28
```

```
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Bacillus licheniformis ATCC 14580
US-09-110-959A-2

Query Match      8.4%; Score 143.5; DB 3; Length 261;
Best Local Similarity 23.9%; Pred. No. 1.7e-06;
Matches 63; Conservative 37; Mismatches 97; Indels 67; Gaps 15;

QY 66 PIDKDGNGNPEFYIENLNILNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRSNW---- 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 35 PSDKLYPKNKYYIFNNVWGAQOVSGWQTIYHNSDSDMGWVW-----NWPSNT 83

QY 122 --VHGYPEIFYGNKPNANYATDGPILPSPKSVNLTDFVLTISYKLEPKNGLPINFAIES 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 STVKAYPSIVSG--WHWTGTYTAGS--GFTRLSDOKNINTRKVSIS-ANG-TYNAAYDI 138

QY 180 WL---TREAWRTTGINSDEQEVMIWIYDGLQPGAGSKYKEIVVPIIVNGTPVNAFVEMK 236
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 WLHNTNKASWD---SAPTDEIMIMLNNTNAGPAGSYVETVSIG-----GHSWKVVK 186

QY 237 ANI-----GWYVAFRIKTIKEGTVTIPYGAFISVAANISSLPNY-----TELY 281
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 GYIDAGGKGWNVFSG-IRTANTQ-----SANLIRDFNTYLADSKQWLKSTKY 234

QY 282 LEDVEIGTE-FGTPSTTSAHLEW 304
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 235 VSSVEFGTEVFG--GTGQINISNW 256

RESULT 6
US-09-216-295-23
; Sequence 23, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328el Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Rhodothermus marinus
US-09-216-295-23

Query Match      7.5%; Score 128; DB 3; Length 260;
Best Local Similarity 24.1%; Pred. No. 6.7e-05;
Matches 61; Conservative 32; Mismatches 110; Indels 50; Gaps 11;

QY 59 DGEWPGAPIDKDGNGNPEFYIEI-----NLNLTNATGFAEMTYNLTSGVLHYV 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19 DMLFPDGDNGKEPEPEPEPTVELCGRWDAWDVAGGRYRVNNVWGAETAQCIEVGL--- 74

QY 108 QQLDNIVL-----RDRSNVHGYPEIFYGNKPNANYATDGPILPSPKSVNLTDFVLTISY 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 75 -ETGNFTITRADHDNGNNVAAYPAYVFGCH-WAPARAIRDCARAGAVRRRAHELDVT--- 129

QY 164 KLEPKNGLP-----NFAIESWLTREAWRTTGINSDEQEVMIWIYD-GLQPGAGSKYKEI 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 -----PITGRNNAAYDIWFSPVTNSGNGY-SGGAELMIWLWNGGWMGPGGRVATV 180

QY 218 VVPIIVNGTPVNAFVEMKANKICWEYVAFRIKTIKEGTVTIPYGAFISVAANISSLPNY 277
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 EL-----AGATWEVYVADWDMNVIAVRRITP-TTSVSELDLAKFIDDAVARGVI--R 229

QY 278 TELYLEDVEIGTE 290
```


Db 374 TTTTSTPPTTTTSTPTTTTQTPTTTTPTTTTTPSNNVPFVNV-LPTSSQ 432
QY 62 WPGAPDKDGN---PEFYIENLWNIL---NAT-----GFAEMT 96
Db 433 YEGTSVEVCDGTQACSSVWGAPNLGVVKGIGNATMDPNVGMWEDVYKTAPODGTGSK 492
QY 97 YNLTSGLVHYVQQLDNIIVLDRSNVHGYPFIFGKPKWNAVATDGP-IPLPSKVSNL 154
Db 493 MEIRNGVLK-VTNLWNINHPKNTM-AYPEVIYGAKPW-GNQPINAPNFVLPKVSQ 548

RESULT 10

US-09-066-544-2
; Sequence 2, Application US/09066544
; Patent No. 6001984
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,544
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,572
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-09-066-544-2

Query Match 7.0%; Score 119.5; DB 3; Length 553;
Best Local Similarity 25.1%; Pred. No. 0.0017;
Matches 45; Conservative 22; Mismatches 51; Indels 61; Gaps 9;

QY 33 STNTSSTPPTTLSTT-----KVLKIRYPDGE 61
Db 374 TTTTSTPPTTTTSTPTTTTQTPTTTTPTTTTTPSNNVPFVNV-LPTSSQ 432
QY 62 WPGAPDKDGN---PEFYIENLWNIL---NAT-----GFAEMT 96
Db 433 YEGTSVEVCDGTQACSSVWGAPNLGVVKGIGNATMDPNVGMWEDVYKTAPODGTGSK 492
QY 97 YNLTSGLVHYVQQLDNIIVLDRSNVHGYPFIFGKPKWNAVATDGP-IPLPSKVSNL 154
Db 493 MEIRNGVLK-VTNLWNINHPKNTM-AYPEVIYGAKPW-GNQPINAPNFVLPKVSQ 548

RESULT 11

US-08-951-086-2
; Sequence 2, Application US/08951086
; Patent No. 6074867
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,086
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/651,572
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-951-086-2

Query Match 7.0%; Score 119.5; DB 3; Length 553;
Best Local Similarity 25.1%; Pred. No. 0.0017;
Matches 45; Conservative 22; Mismatches 51; Indels 61; Gaps 9;

QY 33 STNTSSTPPTTLSTT-----KVLKIRYPDGE 61
Db 374 TTTTSTPPTTTTSTPTTTTQTPTTTTPTTTTTPSNNVPFVNV-LPTSSQ 432
QY 62 WPGAPDKDGN---PEFYIENLWNIL---NAT-----GFAEMT 96
Db 433 YEGTSVEVCDGTQACSSVWGAPNLGVVKGIGNATMDPNVGMWEDVYKTAPODGTGSK 492
QY 97 YNLTSGLVHYVQQLDNIIVLDRSNVHGYPFIFGKPKWNAVATDGP-IPLPSKVSNL 154
Db 493 MEIRNGVLK-VTNLWNINHPKNTM-AYPEVIYGAKPW-GNQPINAPNFVLPKVSQ 548

RESULT 12

US-09-430-669-2
; Sequence 2, Application US/09430669
; Patent No. 6329187
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND


```

; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptomyces lividans CelB
US-09-632-575-52

Query Match      6.7%; Score 114.5; DB 4; Length 381;
Best Local Similarity 27.1%; Pred. No. 0.003;
Matches 52; Conservative 21; Mismatches 80; Indels 39; Caps 10;

QY 125 YPELFYGNKPNANYATDGPPLPSKVSNLTDFTLTISYKLEPKNGLPINFAIESWLTRE 184
Db 97 YPSVFNGCHYTNCSFGTD----LPVRLDTVSAAPSSISYGF--VDGAVYNASYDIWLDPT 150

QY 185 AWRRTGINSDEQEVMIWYYDG-LQPAGSKYKEIWPFIIVNGTPV--NATFEVM----- 235
Db 151 A-RTDGVN--QTEIMWFNRVGPQPICGSPV-----GTASVGGRTWEVMSGNGS 197

QY 236 -----KANIGWEY-VAPRIKTIKEGTVTIPYGAFISVAANISSLENYTELYEDV 285
Db 198 NDVLSFVAPSAISGWSFDMDFVRATVARGLAENDW-YLTSVQAGPEPWONGAGLAVNSF 256

QY 286 EIGTEFGTPTST 297
Db 257 SSTVETGTPGCT 268
```

Search completed: June 29, 2004, 20:14:10
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2004, 10:56:46 ; Search time 603 Seconds
(without alignments)
6763.300 Million cell updates/sec

Title: US-09-914-543-45
Perfect score: 960
Sequence: 1 atgagcaagaaaagttcgt.....atagacctttatttcttaa 960

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseqn_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	960	100.0	960	2 AAT94215	Aat94215 Pyrococcus
2	960	100.0	960	2 AAV47540	Aav47540 Pyrococcus
3	960	100.0	960	2 AAV36924	Aav36924 Pyrococcus
4	960	100.0	960	9 ADC26951	Adc26951 Pyrococcus
5	958.4	99.8	1134	6 ABA91905	Abag1905 Pyrococcus
6	958.4	99.8	1134	6 ABA0497	Abag0497 P. furios
7	536.8	55.9	978	7 ACC44577	Acc44577 Pyrococcus
8	534	55.6	903	7 ACC44577	Acc44577 pNOV4800
9	115.8	12.1	2529	2 AAT94193	Aat94193 Archaeobac
10	114.2	11.9	894	2 AAV40651	Aav40651 Endo-1,4-
11	114.2	11.9	894	2 AAV47544	Aav47544 pMB447A p
12	47.4	4.9	2000	7 ADA71938	Ada71938 Rice gene
13	47.2	4.9	999	2 AAX80187	Aax80187 Sulfolobu
14	44.2	4.6	999	2 AAX80188	Aax80188 Sulfolobu
15	43.6	4.5	53585	2 AAX20251	Aax20251 Borrelia
16	38.6	4.0	2150	6 ABA85746	Abk85746 DNA encod
17	37.8	3.9	728	2 AAV74922	Aav74922 Staphyloc
18	37.8	3.9	1341	7 ACF74977	Acf74977 Staphyloc
19	37.2	3.9	2338	6 ABA63651	Abk63651 Rat sequ
20	37.2	3.9	2338	6 ABA58086	Abd58086 Toxicity-
21	37.2	3.9	2338	9 ABA52565	Abv52565 Primary r
22	36.6	3.8	469	5 ABA52725	Abv52725 Human pr
23	36.6	3.8	110000	6 ABA90521_21	Continuation (22 o

24	36.4	3.8	741	2 AAX98933	Aax98933 Human val
25	36.4	3.8	25580	2 AAX12956	Aax12956 Enterococ
26	36.4	3.8	25580	6 ABA98751	Abv98751 Enterococ
27	36.2	3.8	2000	7 ADA71938	Ada71938 Rice gene
28	36	3.8	936	4 AAF75659	Aaf75659 Human ztr
29	36	3.8	936	8 ACD25921	Acd25921 Degenerat
30	36	3.8	936	9 ADC17391	Adc17391 Human ear
31	36	3.8	12393	6 ABL33263	Ab133263 Human imm
32	35.6	3.7	497	6 ABA45944	Abk45944 CDNA enco
33	35.6	3.7	506	6 ABA87188	Abv87188 Human col
34	35.6	3.7	1083	7 ACA29166	Aca29166 Prokaryot
35	35.4	3.7	2568	7 AAX84329	Aax84329 Stealth v
36	35.4	3.7	26747	6 AA167784	Aa167784 Nucleotid
37	35.2	3.7	496	5 AAH81653	Aah81653 Human dif
38	35.2	3.7	1492	5 AAS86756	Aas86756 DNA encod
39	35.2	3.7	1548	4 AA159507	Aa159507 Human pol
40	35	3.6	14041	4 AAH48024	Aah48024 Internal
41	34.8	3.6	3146	4 AA158397	Aa158397 Human pol
42	34.8	3.6	3146	8 ABA849367	Abd849367 Novel hum
43	34.8	3.6	3150	6 ABL65914	Ab165914 Lung canc
44	34.8	3.6	3150	6 ABL69587	Ab169587 Prostate
45	34.8	3.6	3150	6 ABL64453	Ab164453 Stomach c

ALIGNMENTS

RESULT 1
AAT94215
ID AAT94215 standard; DNA; 960 BP.
XX
AC AAT94215;
XX
AC AAT94215;
DT 17-OCT-2003 (revised)
DT 21-MAY-1998 (first entry)
XX
XX Pyrococcus furiosus endoglucanase DNA.
XX
XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;
KW beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW thermostable enzyme; thermophilic; glycosidase; ss.
XX
XX Pyrococcus furiosus; (clone 7EG1).
XX
XX WO9744361-A1.
XX
PD 27-NOV-1997.
XX
XX
PF 22-MAY-1997; 97WO-US008793.
XX
XX 22-MAY-1996; 96US-00651572.
(RECO-) RECOMBINANT BIOTECNOLOGY INC.
PI Lam DE, Mathur EJ;
XX
XX WPI; 1998-018435/02.
DR P-PSDB; AAW35007.
XX
XX Endo:glucanase(s), preferably form archaeal bacterium, AEPII 1a - useful
PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic
PT bonds in cellulose.
XX
XX Claim 3; Fig 1W; 164pp; English.
PS
XX This DNA sequence from Pyrococcus furiosus (clone 7EG1) encodes an
CC endoglucanase (see AAW35007) that is able to degrade
CC carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic bonds in
CC cellulose, and which shows homology to the thermostable endoglucanase
CC (see AAW34985) of archaeobacterium hydrothermal vent isolate AEPIIIa. The
CC DNA can be used in the recombinant production of the endoglucanase and as
CC a probe to identify similar sequences. 24 Endoglucanase polynucleotides
CC (see AAT94193-216) are claimed. These can be incorporated into plasmid or

QY 61 TATTTTGTAGAAAAGTATCATACCTCTGAGGACAAAGTCAACTTCAAAATACCTCATCTACA 120
 Db |||||
 QY 61 TATTTTGTAGAAAAGTATCATACCTCTGAGGACAAAGTCAACTTCAAAATACCTCATCTACA 120
 Db |||||
 QY 121 CCACCCCAAAACAACTTTCACTACCAAGGTTCTCAAGATTAGATACCTGTAGCGGT 180
 Db |||||
 QY 121 CCACCCCAAAACAACTTTCACTACCAAGGTTCTCAAGATTAGATACCTGTAGCGGT 180
 Db |||||
 QY 181 GAGTGGCCAGGAGCTCTTATTGATAAGGATGGTATGGAAACCCAGAAATCTCATATTGAA 240
 Db |||||
 QY 181 GAGTGGCCAGGAGCTCTTATTGATAAGGATGGTATGGAAACCCAGAAATCTCATATTGAA 240
 Db |||||
 QY 241 ATAAACCTATGGAACATTTCTTAATGCTACTGGAATTTGCTGAGATGACGTACAAATTAACC 300
 Db |||||
 QY 241 ATAAACCTATGGAACATTTCTTAATGCTACTGGAATTTGCTGAGATGACGTACAAATTAACC 300
 Db |||||
 QY 301 AGCGGCTCTTCACTACGTCACAACTTCCACACATGCTTGGAGGATAGAAATTAAT 360
 Db |||||
 QY 301 AGCGGCTCTTCACTACGTCACAACTTCCACACATGCTTGGAGGATAGAAATTAAT 360
 Db |||||
 QY 361 TGGGTGCAATGATACCCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
 Db |||||
 QY 361 TGGGTGCAATGATACCCCGAAATATTTCTATGGAACCAAGCCATGGAATGCAAACTACGCA 420
 Db |||||
 QY 421 ACTGATGCCCCAATACCAATTAACCAAGTAAAGTTTCAAACTTAACAGACTTCTATTAACA 480
 Db |||||
 QY 421 ACTGATGCCCCAATACCAATTAACCAAGTAAAGTTTCAAACTTAACAGACTTCTATTAACA 480
 Db |||||
 QY 481 ATCTCTATAAATCTTGACCCCAAGAAACCGCTGCAATTAACCTTCAATAGAAATCTGCG 540
 Db |||||
 QY 481 ATCTCTATAAATCTTGACCCCAAGAAACCGCTGCAATTAACCTTCAATAGAAATCTGCG 540
 Db |||||
 QY 541 TTAACGAGAGAAGCTTGAGAACCAACAGGAATTAACGCGATGAGCAAGTAATGATA 600
 Db |||||
 QY 541 TTAACGAGAGAAGCTTGAGAACCAACAGGAATTAACGCGATGAGCAAGTAATGATA 600
 Db |||||
 QY 601 TGGATTTACTATGAGGATTAACACCGCTGCTCCAAAGTTAAGGAGATGAGTCCCA 660
 Db |||||
 QY 601 TGGATTTACTATGAGGATTAACACCGCTGCTCCAAAGTTAAGGAGATGAGTCCCA 660
 Db |||||
 QY 661 ATAATAGTAAACGGAACACCACTAAATGCTACATTTGAAATGATGGAAGGCAAACTGGT 720
 Db |||||
 QY 661 ATAATAGTAAACGGAACACCACTAAATGCTACATTTGAAATGATGGAAGGCAAACTGGT 720
 Db |||||
 QY 721 TGGAGATGTTGCTATTAGATTAAGACCCCAATCAAGAGGAAACAGTCAAAATTTCCA 780
 Db |||||
 QY 721 TGGAGATGTTGCTATTAGATTAAGACCCCAATCAAGAGGAAACAGTCAAAATTTCCA 780
 Db |||||
 QY 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTTCAAGCTTACCAAAATTAACAGAACTT 840
 Db |||||
 QY 781 TACGAGCATTTATAAGTGTTCAGCCCAACATTTTCAAGCTTACCAAAATTAACAGAACTT 840
 Db |||||
 QY 841 TACTTAGAGGACGTGGAGATGGAACCTGAGTTTGGAAACCGCAAGCACTACCTCGCCAC 900
 Db |||||
 QY 841 TACTTAGAGGACGTGGAGATGGAACCTGAGTTTGGAAACCGCAAGCACTACCTCGCCAC 900
 Db |||||
 QY 901 CTAGATGGTGGATCACAAACATACCACTAACTCTCTAGATAGACCTCTTATTTCTAA 960
 Db |||||
 QY 901 CTAGATGGTGGATCACAAACATACCACTAACTCTCTAGATAGACCTCTTATTTCTAA 960
 Db |||||

RESULT 3

AAV36924

ID AAV36924 standard; DNA; 960 BP.

XX AC AAV36924;

XX XX

DT 17-OCT-2003 (revised)

DT 21-DEC-1998 (first entry)

XX XX

DE Pyrococcus VC1-7EG1 glycosidase gene coding region.

XX Glycosidase; VC1-7EG1; thermostable enzyme; oligosaccharide; glucose;

KW sugar; baking; textile; detergent; beta-galactosidase; ss.
 XX Pyrococcus furiosus; strain VC1-7EG1.
 OS
 PN MO9824799-A1.
 XX
 XX PD 11-JUN-1998.
 XX
 XX PF 08-DEC-1997; 97WO-US022623.
 XX
 XX PR 06-DEC-1996; 96US-0056916P.
 PR 10-OCT-1997; 97US-00949026.
 XX
 XX PA (DIVE-) DIVERSA CORP.
 XX
 PI Bylina EJ, Swanson RV, Mathur EJ, Lam DE;
 XX
 XX WPI; 1998-352407/31.
 DR P-PSDB; AAW49875.
 XX

Glycosidase enzymes from organisms of the genera Staphylothermus, Pyrococcus and Thermococcus - for deriving sugar from oligosaccharides, useful in the e.g. food processing, textile or baking industries.

Claim 1; Fig 18a-b; 92pp; English.

This isolated polynucleotide comprises a coding region for glycosidase VC1-7EG1 (see AAW49875) from a clone (7EG1) of Pyrococcus furiosus VC1, which grows optimally at 100 degC. The invention provides 18 polynucleotides (see AAV36907-24) coding for thermostable glycosidases (see AAW49558-75) having glucosidase, alpha-galactosidase, beta-galactosidase, beta-mannosidase, beta-mannanase, endoglucanase or pullulanase activity. Vectors and host cells are also claimed. A method is provided for producing the enzymes by recombinant techniques. A claimed method for generating glucose from soluble cell oligosaccharides comprises contacting a sample (selected from dairy products, fruit juice, detergent, textile, guar gum, animal feed, plant biomass or waste product) containing oligosaccharides (selected from maltose, cellobiose, lactose, sucrose, raffinose, stachyose, verbascose, cellobiose, starch, amylose, glycogen, disaccharides, polysaccharides and pullulan) with one of the claimed glycosidases such that glucose is produced. (Updated on 17 -OCT-2003 to standardise OS field)

SQ Sequence 960 BP; 329 A; 207 C; 186 G; 238 T; 0 U; 0 Other;

Query Match 100.0%; Score 960; DB 2; Length 960;

Best Local Similarity 100.0%; Pred. No. 1.4e-276;

Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCAAGAAAAGTTGCTCATCTATCTTAAACAATCCTTTTAGTACAGGCAATA 60

Db |||||

1 ATGAGCAAGAAAAGTTGCTCATCTATCTTAAACAATCCTTTTAGTACAGGCAATA 60

QY 61 TATTTGTAGAAAAGTATCATACCTCTGAGGACAAAGTCAACTTCAAAATACCTCATCTACA 120

Db |||||

61 TATTTGTAGAAAAGTATCATACCTCTGAGGACAAAGTCAACTTCAAAATACCTCATCTACA 120

QY 121 CCACCCCAAAACAACTTTCACTACCAAGGTTCTCAAGATTAGATACCTGTAGCGGT 180

Db |||||

121 CCACCCCAAAACAACTTTCACTACCAAGGTTCTCAAGATTAGATACCTGTAGCGGT 180

QY 181 GAGTGGCCAGGAGCTCTTATTGATAAGGATGGTATGGAAACCCAGAAATCTCATATTGAA 240

Db |||||

181 GAGTGGCCAGGAGCTCTTATTGATAAGGATGGTATGGAAACCCAGAAATCTCATATTGAA 240

QY 241 ATAAACCTATGGAACATTTCTTAATGCTACTGGAATTTGCTGAGATGACGTACAAATTAACC 300

Db |||||

241 ATAAACCTATGGAACATTTCTTAATGCTACTGGAATTTGCTGAGATGACGTACAAATTAACC 300

QY 301 AGCGGCTCTTCACTACGTCACAACTTGGACCAATTTGCTGAGGATAGAAATTAAT 360

Db |||||

301 AGCGGCTCTTCACTACGTCACAACTTGGACCAATTTGCTGAGGATAGAAATTAAT 360

QY 661 ATAATAGTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATGGAAGCAAAACATGGT 720
 Db |||||||
 QY 661 ATAATAGTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATGGAAGCAAAACATGGT 720
 Db |||||||
 QY 721 TGGGAGTATGTTGCAATTTAGATAAAGAACCCCAATCAAAAGAGGGAACAGTGACAATTCGA 780
 Db |||||||
 QY 721 TGGGAGTATGTTGCAATTTAGATAAAGAACCCCAATCAAAAGAGGGAACAGTGACAATTCGA 780
 Db |||||||
 QY 781 TACGGAGCATTTATAAGTGTGTCAGCCCAACATTTCAAGCTTTACCAATTTACACAGAATCT 840
 Db |||||||
 QY 781 TACGGAGCATTTATAAGTGTGTCAGCCCAACATTTCAAGCTTTACCAATTTACACAGAATCT 840
 Db |||||||
 QY 841 TACTTAGAGGAGCTGGAGATTGGAACCTGAGTTTGAAGCCCAAGCACTACCTCGGCCAC 900
 Db |||||||
 QY 841 TACTTAGAGGAGCTGGAGATTGGAACCTGAGTTTGAAGCCCAAGCACTACCTCGGCCAC 900
 Db |||||||
 QY 901 CTAGAGTGGTGATCACAAACATACACTAACTCTCTAGATAGACCTCTATTTCCTAA 960
 Db |||||||
 QY 901 CTAGAGTGGTGATCACAAACATACACTAACTCTCTAGATAGACCTCTATTTCCTAA 960
 Db |||||||

RESULT 5

ABA91905

ID ABA91905 standard; DNA; 1134 BP.

XX ABA91905;

AC ABA91905;

XX ABA91905;

DT 15-MAY-2002 (first entry)

XX 15-MAY-2002 (first entry)

XX Pyrococcus furiosus thermostable cellulase egIA gene.

XX Cellulase; EgIA; thermostable; enzyme; gene; ds.

XX Pyrococcus furiosus.

XX Pyrococcus furiosus.

XX Key Location/Qualifiers

FH 82..1041

FT /*tag= a

FT /product= "Cellulase_EgIA"

FT sig_peptide 82..138

FT /*tag= b

FT mat_peptide 139..1038

FT /*tag= c

FT WO200196382-A2.

XX 20-DEC-2001.

XX 15-JUN-2001; 2001WO-ISO000012.

XX 15-JUN-2000; 2000US-00594884.

XX (PROK-) PROKARIA EHF.

XX Wicher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;

XX WPI; 2002-226850/28.

XX P-PSDB; AAM50997.

XX Novel thermostable variant cellulase which is truncated such that amino

XX terminal hydrophobic region and linker group of corresponding full length

XX enzyme is deleted, has improved catalytic properties and/or stability.

XX Disclosure; Page 45-47; 51pp; English.

XX The present sequence is that of the Pyrococcus furiosus egIA gene

XX encoding thermostable cellulase EgIA (see AAM50997), a family 12 glycosyl

XX hydrolase. The invention provides polypeptides having thermostable

XX cellulase activity. These are variants of a glycosyl hydrolase of family

XX 12 and are truncated such that 1 or more of the amino acid residues

XX corresponding to position 1 to about 40 are deleted. The polypeptides are

XX particularly derived from thermophilic Rhodothermus and Pyrococcus spp.

XX For EgIA, amino acids from the N-terminal hydrophobic region and/or the

CC linking moiety are deleted. Claimed polypeptides (see AAM50976-996) are
 CC also derived from Rhodothermus marinus Celi2A cellulase. Isolated nucleic
 CC acids encoding these thermostable cellulase polypeptides, and host cells,
 CC are also claimed. The thermostable variant cellulases are useful e.g. in
 CC the wood and paper pulp industries, in detergent compositions, in the
 CC textile industry, to improve the feed value of animal feedstuffs, and in
 CC the food industry
 XX

SQ Sequence 1134 BP; 392 A; 228 C; 223 G; 291 T; 0 U; 0 Other;

Query Match 99.8%; Score 958.4; DB 6; Length 1134;

Best Local Similarity 99.9%; Pred. No. 4.5e-276;

Matches 959; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAGCAAGAAAAGTTGCTCATCTCTATCTTAAACAATCTCTTTTAGTACAGGCAATA 60
 Db |||||||
 QY 82 ATGAGCAAGAAAAGTTGCTCATCTCTATCTTAAACAATCTCTTTTAGTACAGGCAATA 141
 Db |||||||
 QY 61 TATTTGTAGAAAAGTATCATACCTCTGAGGCAAGTCAACTTCAAAATACCTCATCTACA 120
 Db |||||||
 QY 142 TATTTGTAGAAAAGTATCATACCTCTGAGGCAAGTCAACTTCAAAATACCTCATCTACA 201
 Db |||||||
 QY 121 CCACCCCAACACACACTTTCCCTACCAAGGTTCTCAAGATTAGATACCTCATGACGGT 180
 Db |||||||
 QY 202 CCACCCCAACACACACTTTCCCTACCAAGGTTCTCAAGATTAGATACCTCATGACGGT 261
 Db |||||||
 QY 181 GAGTGCCAGGAGCTCTCTATTGATAGGATGGTGGGAACCCAGAAATTTACATTGAA 240
 Db |||||||
 QY 262 GAGTGCCAGGAGCTCTCTATTGATAGGATGGTGGGAACCCAGAAATTTACATTGAA 321
 Db |||||||
 QY 241 ATAACTATGGAACATCTTAACTGCTACTGATTTGCTGAGATGACGTAACATTAAAC 300
 Db |||||||
 QY 322 ATAACTATGGAACATCTTAACTGCTACTGATTTGCTGAGATGACGTAACATTAAAC 381
 Db |||||||
 QY 301 AGCGGGCTCTTCACTACGTCACCAACACTTGACACATTTGCTTGAGGATAGAGTAAT 360
 Db |||||||
 QY 382 AGCGGGCTCTTCACTACGTCACCAACACTTGACACATTTGCTTGAGGATAGAGTAAT 441
 Db |||||||
 QY 361 TGGTGTCATGATACCCCGAATAATTTCTATGGAACAAGCCATGGAATGCAAACTACGCA 420
 Db |||||||
 QY 442 TGGTGTCATGATACCCCGAATAATTTCTATGGAACAAGCCATGGAATGCAAACTACGCA 501
 Db |||||||
 QY 421 ACTGATGCCCCAATACCATTTACCCAGTAAAGTTTCAAACTCAACAGACTTCTATCTACA 480
 Db |||||||
 QY 502 ACTGATGCCCCAATACCATTTACCCAGTAAAGTTTCAAACTCAACAGACTTCTATCTACA 561
 Db |||||||
 QY 481 ATCTCTATAAACTTTGAGCCCCAAGAACCGGCTGCCAATTTAACTTCGCAATAGAACTCTGG 540
 Db |||||||
 QY 562 ATCTCTATAAACTTTGAGCCCCAAGAACCGGCTGCCAATTTAACTTCGCAATAGAACTCTGG 621
 Db |||||||
 QY 541 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAGAGTAATGATA 600
 Db |||||||
 QY 622 TTAACGAGAGAGCTTGGAGAACACAGGAATTAACAGCGATGAGCAGAGTAATGATA 681
 Db |||||||
 QY 601 TGGATTACTATGACGGATTACAACCGGCTGGCTCCAAAGTTTAAAGAGATTCTAGTCCCA 660
 Db |||||||
 QY 682 TGGATTACTATGACGGATTACAACCGGCTGGCTCCAAAGTTTAAAGAGATTCTAGTCCCA 741
 Db |||||||
 QY 661 ATAAATGTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATGGAAGCAAACTGGT 720
 Db |||||||
 QY 742 ATAAATGTTAAACGGAACACCAAGTAAATGCTACATTTGAAGTATGGAAGCAAACTGGT 801
 Db |||||||
 QY 721 TGGGAGTATGTTGCAATTTAGATAAAGAACCCCAATCAAAAGAGGGAACAGTGACAATTCGA 780
 Db |||||||
 QY 802 TGGGAGTATGTTGCAATTTAGATAAAGAACCCCAATCAAAAGAGGGAACAGTGACAATTCGA 861
 Db |||||||
 QY 781 TACGGAGCATTTATAAGTGTGTCAGCCCAACATTTCAAGCTTTACCAATTTACACAGAATCT 840
 Db |||||||
 QY 862 TACGGAGCATTTATAAGTGTGTCAGCCCAACATTTCAAGCTTTACCAATTTACACAGAATCT 921
 Db |||||||
 QY 841 TACTTAGAGGAGCTGGAGATTGGAACCTGAGTTTGAAGCCCAAGCACTACCTCGGCCAC 900
 Db |||||||
 QY 922 TACTTAGAGGAGCTGGAGATTGGAACCTGAGTTTGAAGCCCAAGCACTACCTCGGCCAC 981
 Db |||||||

	901	CTAGAGTGGTGATCACAAACATAACAACACTAACCCTCTGTAGATGAGACCTCTTTATTTCCTAA	960
QY			
Db	982	CTAGAGTGGTGATCACAAACATAACAACACTAACCCTCTGTAGATGAGACCTCTTTATTTCCTAA	1041
RESULT 6			
ABA00497	ID	ABA00497 standard; cDNA; 1134 BP.	
XX	AC	ABA00497;	
XX	DT	07-FEB-2003 (first entry)	
XX	P.	furius EglA coding sequence.	
DE	DE	Gene; variant; thermostable; cellulase; Cell12A; family 12; EglA;	
XX	KW	glycosyl hydrolase; freeness; ink; coating; toner; colour; wood;	
KW	KW	pulp; detergent; cellulose-containing textile; garment; lint;	
KW	KW	fibrous crop; fruit; vegetable; grain; feed value; stability; solubility;	
KW	KW	catalytic activity; cytotoxicity; ss.	
OS	Pyrococcus furiosus.		
XX	Key	Location/Qualifiers	
FH	CDS	82..1041	
FT	/tag=	a	
FT	/product=	"EglA"	
XX	US2002102699-A1.		
PN	XX		
PD	01-AUG-2002.		
XX	23-OCT-2001; 2001US-00003759.		
PF	15-JUN-2000; 2000US-00594884.		
PR	15-JUN-2001; 2001WO-IS000012.		
PR	(PROK-) PROKARIA LTD.		
PA	Wiher KB, Holst OP, Hachem MYA, Karlsson EMN, Hreggvidsson GO;		
PI	WPI: 2002-749592/81.		
DR	P-PSDB; AAG79624.		
XX	Novel isolated nucleic acid encoding a polypeptide having thermostable		
PT	cellulase activity, useful for producing thermostable cellulase		
PT	polypeptide and as probes for isolating homologous sequences.		
XX	Disclosure; Page 14-15; 21pp; English.		
PS	This sequence encodes EglA, a variant family 12 glycosyl hydrolase		
XX	derived from P. furiosus. The amino terminal of EglA, amino acids 1-19,		
CC	constitute the hydrophobic domain. Residues 28-49 constitute the linker		
CC	molecule, with amino acids 50-319 forming the catalytic domain. EglA is		
CC	useful to improve freeness and to remove inks, coatings, toners and		
CC	colours from wood or paper pulp, in detergent compositions and to treat		
CC	cellulose-containing textiles and garments to improve the feel of the		
CC	fabric or to remove lint, in the treatment of fibrous crops, fruits and		
CC	vegetables or grains to improve feed values or to extract starches (e.g.		
CC	sugars) or other components of the crop, fruit, vegetable or grain being		
CC	treated). The variant glycosyl hydrolase polypeptide has improved		
CC	characteristics, such as increased stability (e.g. thermal stability,		
CC	detergent stability), increased solubility in aqueous solvents, increased		
CC	catalytic activity (e.g. specific activity, catalytic rate) and/or		
CC	reduced cytotoxicity relative to the native or full-length thermostable		
CC	cellulase, but retains the substrate specificity of the native or full-		
CC	length cellulase		
XX	Sequence 1134 BP; 392 A; 228 C; 223 G; 291 T; 0 U; 0 Other;		
SQ	Query Match	99.8%; Score 958.4; DB 6; Length 1134;	
	Query Identity	90.9%; Posed No. 4, 5e-276;	

DT 02-JUN-2003 (first entry)
 DE pNOV4800 nucleotide sequence SEQ ID NO:56.
 XX
 KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;
 KW pullulanase; alpha-glucosidase; glucose isomerase; glucoamylase;
 KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;
 KW maltodextrin; ethanol; fermentation; beverage; enzyme; gene; ds.
 XX
 OS Pyrococcus furiosus.
 OS Synthetic.
 XX
 PN WO2003018766-A2.
 XX
 PD 06-MAR-2003.
 XX
 PF 27-AUG-2002; 2002WO-US027129.
 XX
 PR 27-AUG-2001; 2001US-0315281P.
 XX
 FA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 FI Lanahan MB, Basu SS, Battie CJ, Chen W, Craig J, Kinkema M;
 XX
 DR WPI; 2003-268420/26.
 XX
 PT Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.
 PT alpha-amylase, useful for producing plant to produce food products having
 PT improved taste or fermentable substrates for ethanol.
 XX
 PS Disclosure; Page 114; 158pp; English.
 XX
 CC The present invention describes polynucleotides which encode processing
 CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose
 CC isomerase, or glucoamylase) that are optimised for expression in plants.
 CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic
 CC processing enzymes, which are activated under suitable conditions to act
 CC upon the desired substrate. Also described are self-processing transgenic
 CC plants and plant parts, e.g. grain, which express one or more of these
 CC enzymes and have an altered composition that facilitates plant and grain
 CC processing. Also described is a method (M) for converting starch to
 CC starch-derived products in a transformed plant part (TPP), by activating
 CC the starch processing enzyme contained in it. Transgenic grain is useful
 CC for preparing maltodextrin. A transformed plant (TP) can be used to
 CC produce food products having improved taste and to produce fermentable
 CC substrates for ethanol and fermented beverages. (M) eliminates the need
 CC to mill or physically disrupt the integrity of plant parts prior to
 CC recovery of starch-derived products. The present sequence represents a
 CC pNOV4800 nucleotide sequence (Amy32B signal sequence with EGLA), which is
 CC given in the exemplification of the present invention
 XX
 SQ Sequence 978 BP; 206 A; 371 C; 240 G; 161 T; 0 U; 0 Other;
 Query Match 55.9%; Score 536.8; DB 7; Length 978;
 Best Local Similarity 74.4%; Pred. No. 6.1e-150;
 Matches 676; Conservative 0; Mismatches 232; Indels 0; Gaps 0;
 QY 52 CAGGCAATATATTTGTAGAAAAGTATCATACCTCTGAGGACAGTCAACTTCAAAATACC 111
 Db 70 CATCAATCTACTTCTGTTGAGAGTACACACCTCCGAGGACAAAGTCCACTCCACACC 129
 QY 112 TCATCTACACACCCCAACACTTCCACTACCAAGTTCTCAAGATTAGATACCCCT 171
 Db 130 TCCTCCACCCCGCGAGACACACCTCTCCACCAACCAAGTGTCAAGATCCGCTACCCG 189
 QY 172 GATGACCGTGTGCGGACGAGTCCCTATTGATGAGATGTGATGGACCCAGATTC 231
 Db 190 GACACCGTGTGTTGGCGCGCGCCCGATCGAAGGACGCGCAACCCGGAGTTC 249
 QY 232 TACATCGAGATCAACCTCTGGAACATCTTAAATGCTACTGATTTGCTGAGATGAGGTAC 291
 Db 250 TACATCGAGATCAACCTCTGGAACATCTTCAACGCCACCGGCTTCGCCGAGATGACCTAC 309

292 AATTTAACGAGCGCGCTCTCTCACTACGTCCAACTTGACAACTTGTCTTGAGGAT 351
 |||
 310 AACCTCACTAGTGGCGTGTCTCACTACGTGCGAGCGTCCGACNACATCTGTCTCCGCGAC 369
 |||
 352 AGAAGTAATTTGGTGTGATGGATACCCCGAAATATTCTATGGAAACAAGCCATCGAATGCA 411
 |||
 370 CGCTCCAACTGGGTGACGCGCTACCCGAAATCTTCTACGGCAACAAGCGGTGGACGCGC 429
 |||
 412 AACTAGCAACTGATGGCCCAATACCATTAACCCAGTAAAGTTTCAAAACCTTAACAGACTTC 471
 |||
 430 AACTAGCCACCGACGCGCCCGATCCGCTCCGTCACAGTGTCCAACTCAACCGACTTC 489
 |||
 472 TATCTAACAATCTCTATATACTTGAGCCCAAGACGCGCTGCCAATTAATCTTCGCAATA 531
 |||
 490 TACCTCACCATCTCTCTACAAAGCTCGAGCGCAAGAACGGTCTCCCGATCAACTTCGCGCATC 549
 |||
 532 GAATCTCTGTTAACGAGAGAACTGGAGAACCAAGAGGAATTAACAGCGATCAGCAAGAA 591
 |||
 550 GAGTCTCTGGCTACCCGCGAGCGCTGGCGCACACCGGCATCACTTCCGACGCGAGGAG 609
 |||
 592 GTAATGATATGATTTACTTATGACGGATTACACCGGCTGGCTCCAAAGTTAAGGAGATT 651
 |||
 610 GTGATGATCTGGATCTACTACGAGCGCTCCAGCGCGCGGGCTCCAAAGTGAAGGAGATC 669
 |||
 652 GTAGTCCCAATTAATAGTTAACCGAACAACCGATTAATGCTACATTTGAAGTATGGAGCA 711
 |||
 670 GTGGTCCGATCATCGTGAACGCGACCCCGGTGAACGCGCACCTTCGAGGTGTGGAAGGCC 729
 |||
 712 AACATTGGTTGGAGTATGTTCATTTAGAAATAAAGACCCCAATCAAGAGGAGGAAACAGTG 771
 |||
 730 AACATCGCTGGAGTACGTGGCTTCGCAATCAAGACCCCGATCAAGAGGCGGCACCGTG 789
 |||
 772 ACAATTCATACGGAGCATTTATAGTTGTGAGCAACAATTTCAAGCTTACAAATTAC 831
 |||
 790 ACCATCCCGTACGGCGCTTCTATCTCCGTGGCGGCAACATCTCTCCCTCCCGAACTAC 849
 |||
 832 ACAGAACTTTACTTAGAGACGCTGGAGATTGGAATGAGTTTGGAAACGCCAAGCACTACC 891
 |||
 850 ACGAGAGTACTCTGAGGACCTGGAGATCGGACCGGATTCGGACCCCGCTCCACCACC 909
 |||
 892 TCGGCCCACTAGTGGTGGATCAACAATCAACTAATCTCTTAGATAGACTCTTT 951
 |||
 910 TCGGCCCACTAGTGGTGGATCAACAATCAACTAATCTCTTAGATAGACTCTTT 969
 |||
 952 ATTTCCTA 959
 |||
 970 ATCTCCTA 977
 |||
 RESULT 8
 ACC44577
 ID ACC44577 standard; cDNA; 903 BP.
 XX
 AC ACC44577;
 XX
 DT 02-JUN-2003 (first entry)
 XX
 DE Pyrococcus furiosus EGLA nucleotide sequence SEQ ID NO:56.
 XX
 KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;
 KW pullulanase; alpha-glucosidase; glucose isomerase; glucoamylase;
 KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;
 KW maltodextrin; ethanol; fermentation; beverage; enzyme; gene; ss.
 XX
 OS Pyrococcus furiosus.
 XX
 FH Location/Qualifiers
 CDS 1..903
 FT /*tag= a
 FT /product= "EGLA"
 FT /transl_except= (pos:781..783,aa:1eu)
 FT /note= "no start codon given"
 XX

XX PS Claim 3; Fig 1A; 164pp; English.

XX CC This genomic DNA sequence from archaeal bacterial AEP11a codes for a

CC thermophilic endoglucanase (see AAW34985) that is able to degrade

CC carboxymethylcellulose and to hydrolyse the beta-1,4-glycosidic bonds in

CC cellulose. AEP11a was discovered in a shallow marine hydrothermal vent.

CC It grows optimally at 85 degC and pH 6.5. The DNA (deposited in pOET as

CC ATCC 97516) was isolated from a genomic DNA library of AEP11a by

CC screening for clones having endoglucanase activity. It can be used in the

CC recombinant production of the endoglucanase and as a probe to identify

CC similar sequences. Other polynucleotides (see AAW34194-216) encoding

CC endoglucanases (see AAW34986-W3508) having homology to the AEP11a

CC endoglucanase are also claimed. The endoglucanases can also be used to

CC degrade cellulose for the conversion of plant biomass into fuels and

CC chemicals, for use in detergents, textiles, animal feed, waste treatment,

CC and in the fruit juice and brewing industries for the clarification and

CC extraction of juices. (Updated on 27-AUG-2003 to correct OS field.)

XX SQ Sequence 2529 BP; 702 A; 719 C; 638 G; 470 T; 0 U; 0 Other;

Query Match 12.1%; Score 115.8; DB 2; Length 2529;

Best Local Similarity 66.8%; Pred. No. 8.8e-24;

Matches 165; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 92 ACAAGTCACTTCAATACCTCATCTACACACCCCAACACACATTTCCACTACCAAGG 151

DB 2276 AGACTTCAACCACTACCAACCACTCACCAGCGGACACACCGCACTGCTCAGGACG 2335

QY 152 TTCTCAAGATTAGATACCTGATGACGGTGAGTGCCGAGGAGTCTTATTGATAAGGATG 211

DB 2336 TAATTAGCTCAGTGACCGGACGATGGCGAGTGCCCGAGGCCCAATTTGACAGGATG 2395

QY 212 GTGATGGGAACCCAGAAATTTACATTTGAATAAATACCTATGGAACATTTCTTAATGCTACTG 271

DB 2396 GAGACGGAAACCCAGAGTTCTACATAGAAATAAACCCTGGGACATCTAGCGGCTGAAA 2455

QY 272 GATTGCTGAGATGACGATGATTTAACCGGCGGCTTCTACTAGTCCCAACAACTTG 331

DB 2456 GCTACGGCGGAGATGACCTTACAACTTGAGCAGCGGGGTTCTCCACTAGTCCAGGCCCTGG 2515

QY 332 ACAACAT 338

DB 2516 ATAGTAT 2522

RESULT 10

AAV40651

ID AAV40651 standard; DNA; 894 BP.

XX AC AAV40651;

XX DT 17-OCT-2003 (revised)

XX DT 26-OCT-1998 (first entry)

DE Endo-1,4-beta-glucanase gene in plasmid pMB447A.

XX Endoglucanase; cellulase; thermostable enzyme; textile; biopolishing;

XX stone-washing; saccharification; feedstuff; coffee; ss.

OS Bacillus sp.

OS Dictyoglomus; sp.

OS Chimeric.

XX Key Location/Qualifiers

FT sig_peptide 1..87

FT /tag= a

FT /note= "Bacillus sp. signal peptide"

FT mat_peptide 88..894

FT /tag= b

FT /note= "Dictyoglomus sp. endoglucanase gene"

XX PN W09828410-A1.

XX PD 02-JUL-1998.

XX PF 19-DEC-1997; 97WO-DK000583.

XX PR 20-DEC-1996; 96DK-00001483.

XX PA (NOVO) NOVO-NORDISK AS.

XX PI Schuelein M, Bjornvad ME, Norrevang IA;

XX WPI: 1998-377642/32.

XX P-PSDB; AAW57777.

XX New endo-1,4-beta-glucanase active at high temperature and wide pH range

XX - useful, e.g. for improving properties of cellulosic textiles, e.g.

XX reduce pilling and improve softness.

XX Example 3; Page 32-33; 53pp; English.

XX CC This DNA sequence, from plasmid pMB447A, comprises a coding sequence for

CC a signal peptide (see AAV40648) of a Bacillus sp. and for mature

CC thermostable endoglucanase of Dictyoglomus sp. DSM 6262. The derived

CC protein sequence is provided in AAW57777. The open reading frame is

CC placed under the transcriptional control of the SPAC promoter in pMB447A.

CC The signal peptide portion of the construct directs the endoglucanase to

CC the exterior of host Bacillus sp. cells. Recombinant endoglucanase was

CC expressed in Bacillus subtilis. The enzyme shows optimum activity at a

CC temperature above 85 degC, and exhibits an activity towards CM-cellulose

CC at 70 degC and pH 10 that is higher than 50% relative to the activity at

CC 70 degC and the optimum pH. A Bacillus subtilis strain harboring an

CC expression plasmid encoding the thermostable endoglucanase cloned from

CC Dictyoglomus sp. DSM 6262 is deposited as DSM 11903. The endoglucanase

CC can be used to improve properties of cellulosic textiles, e.g. to reduce

CC pilling, improve softness, to provide a stone-washed look to denim, for

CC industrial cleaning, e.g. of ultrafiltration membranes and pipes, in heat

CC -extrudable polymers (to increase degradability), in conversion of

CC biomass to sugars, in production of ethanol (particularly preliquefaction

CC of grain), for improving digestibility of fodder grains, and in

CC production of instant coffee or other similar extraction processes

CC (increasing capacity of the extraction column) (all claimed). Also

CC provided are DNA constructs encoding the endoglucanase, recombinant

CC expression vectors and host cells (preferably Bacillus, Dictyoglomus or

CC Saccharomyces species or a filamentous fungus). (Updated on 17-OCT-2003

CC to standardise OS field)

XX SQ Sequence 894 BP; 324 A; 167 C; 162 G; 241 T; 0 U; 0 Other;

Query Match 11.9%; Score 114.2; DB 2; Length 894;

Best Local Similarity 51.3%; Pred. No. 1.7e-23;

Matches 291; Conservative 0; Mismatches 273; Indels 3; Gaps 1;

QY 230 TCTACATTGAATAAACCTATGGAACATTTCTTAATGCTACTGATTTGCTGAGATGACCT 289

DB 164 TCACCTTAGAATCACTTCTTGGATATTTGCAATATGCAATATGAGGAATATCATGATGGCAT 223

QY 290 ACAATTTACACGCGCGCTCTCTACGTCCCAACAACTTCACAACTTGCTTGAAGG 349

DB 224 TTATATAAGAGAAGATACCTGTTGAATATTATGCCGACATAAAACATAGTACTTAAGG 283

QY 350 ATAGAAGTAATTGGGTGCATGATACCCGGAATATTCTATGGAACAGCCATGGAATG 409

DB 284 ATAAAAATTCATGGGTACATGATATCTCTGAAGTCTACTATGGGTACAAACCATGGGTG 343

QY 410 CAAACTACGCAACTGATGGCCCAATACCAATTACCCAGTAAAGTTTCAAACTTAACAGACT 469

DB 344 GCCATGGGAATTCATTGAGAAATTAGCTCTTCTTAAAGGTATCAGAAATTCAGAGC 403

QY 470 TCTATCTAACAAATCTCTTATATAAATTTGAGCCCAAGAACCGGCTGCAATTAATCTCGCAA 529

DB 404 TTCTCTTCAATCTAAAAATACACATATGTTAGTACGAGAAATCTTCTCTAAATTTGCTA 463

QY 530 TAGAATCCTGGTTAACGAGAGAAGCTTGGAGAACCAACAGGAATTAACACCGATGAGCAAG 589

Db 464 TGGAAACATGATACAAAGAACCTATCAGAAAACCG---TTACTTTCAGGGGATATAG 520
 Qy 590 AAGTAATGATATGATTTTACTATGACGATTACACCGGCTGCTCCAAAGTTAAGGAGA 649
 Db 521 AGATGATGATGGCTATATGCTATAGACTTTCTCTGCGAGGCGAAGGATGAGGAAG 580
 Qy 650 TTGTAGTCCCAATATAGTTTAAACGGAACACAGTAAATGCTTACATTTTGAAGTATGGAAG 709
 Db 581 TAAATAATACCTATCATCTCTAAACGGTAATCAAAAGACATTATCTGGAAGTATATCTTT 640
 Qy 710 CAACATTTGTTGGGAGTATGCTTATGATTAAGTAAGACCCCAATCAAGAGGACAG 769
 Db 641 CCCCTATGAGCTGGGAGTACGCTGCTATTAATCAAAAGAAATATTTCTTCAAGGACAG 700
 Qy 770 TGACAAATCCATACGAGCAATTTATA 796
 Db 701 TAAATAATACCAATAAATGAATTTTGA 727

RESULT 11
 AAV47544
 ID AAV47544 standard; DNA; 894 BP.
 XX AAV47544;
 XX
 DT 18-NOV-1998 (first entry)
 DE PMB447A plasmid DNA sequence encoding an endo-beta-1, 4-glucanase.
 KW Pyrococcus furiosus endo-beta-1, 4-glucanase; cellulolytic activity;
 KW textile industry; cellulosic fibre; in industrial cleaning process;
 KW sugar; instant coffee; oil industry; hydrocolloid cellulose derivative;
 KW drilling; chimeric; ss.
 XX Synthetic.
 OS Dictyoglomus sp.
 FH Key Location/Qualifiers
 CDS 1..894
 /tag= a
 /product= "Dictyoglomus species endo-beta-1, 4-glucanase"
 sig_peptide 1..87
 /tag= b
 /note= "Signal peptide encoding sequence derived from
 Bacillus species"
 mat_peptide 88..891
 /tag= c
 WO9833895-A1.
 06-AUG-1998.
 30-JAN-1998; 98WO-DK000039.
 31-JAN-1997; 97DK-00000114.
 11-JUL-1997; 97DK-00000853.
 (NOVO) NOVO-NORDISK AS.
 Andersen L, Bjornvad ME, Schuelein M;
 WPI; 1998-437450/37.
 P-PSDB; AAW29730.
 Isolated endo-beta-1,4-glucanase - used for e.g. treating cellulosic
 fibres or polymers, feed production or in oil industry for enhancing oil
 recovery.
 Disclosure; Page 31; 56pp; English.
 The present sequence represents the region of the PMB447A plasmid
 encoding a Dictyoglomus species endo-beta-1, 4-glucanase (EG). This

vector was given as an example of an expression vector that could be used
 in the method of the invention. The invention provides for a Pyrococcus
 furiosus endo-beta-1, 4-glucanase encoding DNA (AAV47540) isolated from a
 P. furiosus genomic DNA library. The invention provides for an enzyme
 composition having EG activity which has optimum activity at a
 temperature of at least 90 deg. C. The EG enzyme exhibits cellulolytic
 activity at extremely high temperatures in a very broad pH range.
 Therefore it is claimed to be useful for, e.g. in the textile industry
 for improving the properties of cellulosic fibres or fabric, for
 providing a stone-washed look of denim; in industrial cleaning processes;
 in the conversion of biomass to sugars; in the production of instant
 coffee or similar extraction processes or in the oil industry for
 degradation of aqueous solutions of hydrocolloid cellulose derivatives
 used in drilling
 XX
 SQ Sequence 894 BP; 324 A; 167 C; 162 G; 241 T; 0 U; 0 Other;
 Query Match 11.9%; Score 114.2; DB 2; Length 894;
 Best Local Similarity 51.3%; Pred No. 1.7e-23;
 Matches 291; Conservative 0; Mismatches 273; Indels 3; Gaps 1;
 Qy 230 TCTACATTGAAATAAACCTATGGAACATCTTTAATGCTACTGGATTCTCTGAGATGACGT 289
 Db 164 TCACCTTAGACTCAACTTTTGGAAATATTGCAAACTATGAAGGAATACATGGATGGCAT 223
 Qy 290 ACAATTTAACAGCGGCTCTTCACTAGCTCCACCACTTCACAACTTGCATCTTCTTGGG 349
 Db 224 TTTATAAGGAAGATATCTGTTGAATATTATGCGGCATATAAACAATAGTACTTAAGG 283
 Qy 350 ATAGAAGTAATTTGGGTGATGATACCCCGAAATATCTTATGGAACAAGCCATGGAATG 409
 Db 284 ATAAAAATTCATGGGTACATGATATCTCTGAAGTCTACTATGGGTACAAACCATGGGCTG 343
 Qy 410 CAACTACGCACTATGCCCCAATACCATACCAGTAAAGTTTCRAAACCCTACAGACT 469
 Db 344 GCCATGGGAATTCCTTATGAGAAATTAGCTCTCTTAAAGAGGTATCAGAAATTCAGAGC 403
 Qy 470 TCTATCTCAACATCTCTATAACTTGAGCCCAAGACCGCTGCCAATTAACCTTCGCA 529
 Db 404 TTCTCTTCATCTAAATACACATATGTACGAGAGAAATCTTCTATAATTTTGCTA 463
 Qy 530 TAGAATCTCTGGTTACGAGAGAGCTTGGAGAACAAACAGGAATTAACAGCGATGAGCAAG 589
 Db 464 TGGAAACATGGAATAACAAAGAACCCCTATCAGAAAACCG---TTACTTTCAGGGGATATAG 520
 Qy 590 AAGTAATGATATGATTTTACTATGACGATTACACCGCTGCTCCAAAGTTAAGGAGA 649
 Db 521 AGATGATGATGGCTATATGCTAATAGACTTTCTCTGCGAGGCGAAGGATGAGGAAG 580
 Qy 650 TTGTAGTCCCAATAATAGTTTAAACGGAACACAGTAAATGCTTACATTTTGAAGTATGGAAG 709
 Db 581 TAAATAATACCTATCATCTCTAAACGGTAATCAAAAGACATTATCTGGAAGTATATCTTT 640
 Qy 710 CAACATTTGTTGGGAGTATGCTTATGATTAAGTAAGACCCCAATCAAGAGGACAG 769
 Db 641 CCCCTATGAGCTGGGAGTACGCTGCTTATAATCAAAAGAAATATTTCTTCAAGGACAG 700
 Qy 770 TGACAAATCCATACGAGCAATTTATA 796
 Db 701 TAAATAATCAATAAATGAATTTTGA 727
 RESULT 12
 ADA71938/c
 ID ADA71938 standard; DNA; 2000 BP.
 XX
 AC ADA71938;
 XX
 DT 20-NOV-2003 (first entry)
 DE
 XX Rice gene, SEQ ID 5263.
 XX Plant; bacterial infection; fungal infection; viral infection; rice;
 KW

Db 27270 AGCAAAATTGGCGCTGATACCAAGAGCCTTTAACCAAGAGCTTTAATGAGCAAGTCAA 27211
Qy 324 ACAACTTGACAAACATTGTCTTGAGGGATAGAGTAATT 361
Db 27210 AGGCTTGGCAGCAGCGGCAAGGGAGATATAAGCAATT 27173

Search completed: July 1, 2004, 21:58:15
Job time : 609 secs

[illegible]

```

RESULT 4
G90291
endoglucanase precursor [imported] - Sulfolobus solfataricus
C/Species: Sulfolobus solfataricus
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C/Accession: G90291
R/She, Q.; Singt, R.K.; Confalonieri, P.; Zivanovic, Y.; Allard, G.; Awaveez, M.
Jong, I.; Jeffries, A.C.; Kozera, C.D.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.;
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A/Description: Sulfolobus solfataricus complete genome.
A/Reference number: A99139
A/Accession: G90291
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-332 <RUR>
A/Cross-References: GB:AE0006641; NID:gl3814564; PIDN:AAK41590.1; GSPDB:GN00155
C/Genetics:
A/Gene: SS01354

```

Qy	4	KKFIYVSLITLLV-----QAIYFVKYHTSDKSTNTSSNP-----PQTL	44
Db	3	KLVIVLPVIVIIAIGVMGGIILYHQOSSLVKPVTITTFSTTSTSTTNAITTVTQVT	62
Qy	47	STTKVLKIRYPDGGWPGAPDKDGDGNPEYIISINLINLNATGPAEMTYNLTSGVLHY	106
Db	63	SITSYNQLIYVTSASSSTPPVYLNNSITPSPYLEVMNNAKYNNGNTVMFNPLRTLSV	122
Qy	107	VQOLDNIVLRDRSNWGHVPEIFYGCKPWNANYATDGPILPSPKVSNLT----DFYLTIS	162
Db	123	SFNLTVQ----KPLEWTNGYPEIYVGRKWDIATYA--GNI-PPMEIGNMTFPWVSFYINLT	176
Qy	163	YKLEPRKGLPINF--AISWLTREAWR---TTGINSDEQEWMIWYIDGLOPAGSKYKEI	217
Db	177	KLDPS-----INFDIASDAWIVRPOIAFSPGTAFGONGDIEIMWLFSQNLQPAGEQGVK	231
Qy	218	WPVLIINGTPVNAATEVMKANI-----GHEYVAFRIKTP-----IKEGVTTPYGAFISVAA	269
Db	232	WPVIYNHILVNATQVWEMKSPVPGWGHEYTAFR---PDGKVINGYVSYEPNLFIKALS	288
Qy	270	NISLSENYTELXLEDVEICTEFTG-PTSTSAHLEWMTNITLTPL	313
Db	289	NFTSY-NITNXYLTDEWFEFTGNTSGNTAYFSVTGNFSETILL	332

Qy 270 NIGSLNPNVTELYLEDVEIKETRGT-PSTTSAAHLEWNIITLPL 313
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 289 NFISY-NIINYILTDWEFGTEGWTSGTAYFSVTVPSETLL 332
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

RESULT 5

H90425

hypothetical protein celB [imported] - Sulfolobus solfataricus

C/Species: Sulfolobus solfataricus

C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C/Accession: H90425

R/Shen, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayero, M.
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A/Description: Sulfolobus solfataricus complete genome.

A/Reference number: A99139

A;Accession: H90425
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-322 <KUR>
A;Cross-references: GB:AE006641; NID:gl3815835; PIDN:AAK42663.1; GSPDB:GN00155
C;Genetics:
A;Gene: celB

Query Match 13.0%; Score 221.5; DB 2; Length 322;
Best Local Similarity 27.3%; Pred. No. 8.1e-10;
Matches 89; Conservative 52; Mismatches 132; Indels 53; Gaps 17;

QY 5 KFTVLSLTLLVQAIIVE--KTH--TSEDKSTNSSTP-----POTLTSTTKVLKIRY 56
DB 3 KLIPFVLGVIIIVLGIIVSIEFGKPHQNASLTRSTERTFLFAHNRPFSL----- 52

QY 57 PDDGEWPGAPDK-----DGDGNPEFYIEINLWNLINATGFAEMTYNLTSGVLHVQOLDN 112
DB 53 ---GNYSNSADALAINLSSNTNATLMVSPFLWNLIGYALGNVNTINI--NYLHVAINLSQ 107

QY 113 IVLDRSRNVHGYPEIFYGNKP--WNAYATDGP--IPLPSKVSNLTDFTLTISYKLEPKN 169
DB 108 I-SKISSNVVDGYPLMGYQBLWPFMYRTTQQLSLPMIVLRLPNFYILNYSVYLIN 166

QY 170 GLPNFALESWLTREAWRTTGINSDEQVMIWYD--GLQAGSKVKKEIIVVPIVNGT 226
DB 167 GSIDDFSYDIWLSNP-NITSLQYGDPEIMLWMNENLSTHTFYIYVGNMSIPTLNGK 225

QY 227 PVNATFEVW-----KANIGWEYVAF--RIKTPKEGTVTIPY-----GAFISVA-ANIS 272
DB 226 IENLSWEVYVLPRTGSAN-GWTGYVFLSPKPEKAEFGVPIGYILKMGVIEKAGWNIY 284

QY 273 SLPNYTELIEDVIGTEGTPPTTS 298
DB 285 NVNTY---YLDIAIQVGEFSDNOQTA 307

RESULT 6
JU0328
cellulase (EC 3.2.1.4) precursor - Erwinia carotovora subsp. carotovora
N;Alternate names: endo-1,4-beta-glucanase; endoglucanase
C;Species: Erwinia carotovora subsp. carotovora
C;Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 08-Oct-1999
C;Accession: JU0328
R;Saarilahti, H.T.; Henrissat, B.; Palva, E.T.
Gene 90, 9-14, 1990
A;Title: CelS: a novel endoglucanase identified from Erwinia carotovora subsp. carotovora
A;Reference number: JU0328; MUID:90337352; PMID:2379837
A;Accession: JU0328
A;Molecule type: DNA
A;Residues: 1-264 <SAA>
A;Cross-references: GB:M32399; NID:gl48389; PIDN:AAA24817.1; PID:gl48390
A;Experimental source: strain SCC3193
A;Note: part of this sequence, including the amino end of the mature protein, was confirmed by N-terminal sequencing
C;Genetics:
A;Gene: celS
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose
A;Pathway: cellulose degradation
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-264/Product: cellulase #status experimental <NAT>

Query Match 6.6%; Score 112.5; DB 2; Length 264;
Best Local Similarity 21.8%; Pred. No. 0.2;
Matches 55; Conservative 38; Mismatches 90; Indels 69; Gaps 14;

QY 68 DKDGDGNGPEFYIEINLWNLINATGFAEMTYNLTSGVLHVQOLDNIVLDRSNW----- 121
DB 40 DKLYFGNNKYVLFNNVNGKDEIKGWQQTIFY-----NSPISMGNWVHWPSSST 86

QY 122 --VHGYPEIFYGNKPWNAYATDGP-IPLPSKVSNLTDFTLTISYKLEPKNGLPINFAIES 179

DB 87 HSKAYPSLVSG-WHWTAGYTENSGLPI--QSSNKSTITSNTVYSIKATG--TYNAAAYDI 141
QY 180 WL---TREAWRTTGINSDEQVMIWYDGLQAGSKVKKEIIVVPIVNGTVPNATFEVWK 236
DB 142 WHHTDKANWD-----SSPTDELMIMLNDTNAGPAGDYIETVFLG-----DSSWVFK 189
QY 237 ANI-----GWEYVAFRIKTPKEGTVTIPYGAFTSVANISSLPNY-----TELY 281
DB 190 GWINADNGGWNVFGP-VHT---SGT-----NSASLNIHRFTDYLVTQKQWMSDEKY 237
QY 282 LEDVEIGTB-FG 292
DB 238 ISSVEFGTIFG 249

RESULT 7
JC2571
cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)
N;Alternate names: endo-1,4-beta-glucanase; endoglucanase
C;Species: Streptomyces rochei
C;Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 22-Oct-1999
C;Accession: JC2571; S34392
R;Perito, B.; Hanhart, E.; Irdani, T.; Iqbal, M.; McCarthy, A.J.; Mastronei, G.
Gene 148, 119-124, 1994
A;Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase
A;Reference number: JC2571; MUID:95011642; PMID:7523249
A;Accession: JC2571
A;Molecule type: DNA
A;Residues: 1-382 <PER>
A;Cross-references: EMBL:X73953; NID:g393391; PIDN:CAAS2139.1; PID:g393392
A;Note: this cellulolytic strain was isolated from the gut of termites
C;Genetics:
A;Gene: eglS
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose
A;Pathway: cellulose degradation
C;Superfamily: bacterial cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-37/Domain: signal sequence #status predicted <SIG>
F;38-382/Product: endoglucanase #status predicted <NAT>
F;279-380/Domain: bacterial cellulose-binding domain homology <BCB>
F;280-379/Disulfide bonds: #status predicted

Query Match 6.1%; Score 103.5; DB 2; Length 382;
Best Local Similarity 25.1%; Pred. No. 1.6;
Matches 43; Conservative 29; Mismatches 70; Indels 29; Gaps 10;

QY 125 YPEIFYGNKPWNAYATDGP-IPLPSKVSNLTDFTLTISYKLEPKNGLPINFAIESWLTR 183
DB 94 YPSVFENG-----CHYTNCSPGTALPARISGISSAPSSISYGF--VDNAVYNASYDIWLDP 146

QY 184 EAWRTTGINSDEQVMIWYDGLQAGSKVKKEIIVVPIVNGT--VNATEVWKANIG 240
DB 147 TP-RDGVN--RTEIMVFNVRVGOIQPGSQV-----GTASVAGRTWEVWGGNG 193

QY 241 W-EYVAFRIKTPKEGTVTIPYGAFTSVANISSLPNYTELILEYDEIGTE 290
DB 194 TNDVLSFVAPSANSSWSPDVMDFVATVARGLAG----NDWYLTSTIQAGFE 240

RESULT 8
T29630
hypothetical protein K09E3.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T29630
R;Johnson, D.; Gattung, S.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid K09E3.
A;Reference number: Z20655
A;Accession: T29630
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA

A;Residues: 1-859 <JOH>
A;Cross-references: EMBL:U41033; PIDN:AAA82378.1; CESP:K09E3.7
C;Genetics:
A;Gene: CESP:K09E3.7
A;Introns: 19/3; 247/3; 428/3; 546/3; 664/3; 782/3

Query Match 6.1%; Score 103; DB 2; Length 859;
Best Local Similarity 21.1%; Pred. No. 5.3;
Matches 67; Conservative 41; Mismatches 115; Indels 94; Gaps 14;

QY 28 TSEDKSTNT-----SSTPPQTTLST-----TKVLKIRYPDDGE 61
DB 336 TLQDQGTNDTNSNRKRAGRSIBELLPDSTITQTSTPITAFISTTTPASTVPTMS 455
QY 62 WFGAPIDXGDGNPEFYIEINLNLNATGFAEMTYNLTSGVLHYVQQLDNIVLRDRS-- 119
DB 456 FPSIP-----HIATLGNVAPINSLTSSLR--IDHQREIPASTLKRRSVS 498
QY 120 -----NWVHGYPEIFYGN-----KP-----WNANYATDGPPLPSKVSNLTFD 157
DB 499 DRMGLPCKRRLNLSVPMSSHVNGSTWEAKPLLPFGWNLHPAMAPFPPTQSTPTSPF 558
QY 158 YLTI-----SY-----KLEPPKGLPINFAIESWLTREAWRTTGINSDEQVMIWI 202
DB 559 IPTTPTTLVPRIMSFPNIHGTISPGNAPINSRVSSLRAMEQQORTVTSASTLKQEA-VFE 617
QY 203 YDGLQ-PAGSKVKEIVVPIVNGTPVNA-----TTEVVKANIGWYVAFRIKTIKEGTVT 258
DB 618 MVNGLSYQRLTNGSGPIPTQFRGNSMEAKOSRPFVWN-----QHPAMPJSLPTQSPPT 672
QY 259 IPYGAFIGVAANISLSP 275
DB 673 IP-AFIPTPSSMP 687

RESULT 9

T20015

hypothetical protein T13H10.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C;Accession: T20015; T24895

R;Coles, L.

submitted to the EMBL Data Library, January 1996

A;Reference number: Z19210

A;Accession: T20015

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-533 <WIL>

A;Cross-references: EMBL:Z68882; PIDN:CAA93112.1; GSPDB:GN00022; CESP:T13H10.1

A;Experimental source: clone C47E12

R;Gajadaty, S.

submitted to the EMBL Data Library, February 1996

A;Reference number: Z19949

A;Accession: T24895

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-533 <W12>

A;Cross-references: EMBL:Z69361; PIDN:CAA93289.1; GSPDB:GN00022; CESP:T13H10.1

A;Experimental source: clone T13H10

C;Genetics:

A;Gene: CESP:T13H10.1

A;Map position: 4

A;Introns: 8/2; 62/3; 101/3; 203/3; 282/2; 355/3; 449/2; 490/3

Query Match

Best Local Similarity 28.7%; Score 102.5; DB 2; Length 533;

Matches 51; Conservative 28; Mismatches 50; Indels 49; Gaps 15;

QY 96 TYNLTSGV-LHYVQQLDNIVLRDRSNWTHGYPEFYGNKPNWANYATDGPPLPSKVSNL 154

DB 227 TMVQAGLGLYHLST-QIHRDTAS-----RNCLYGN-----GQV-----KI 263

QY 155 TDFYLT---ISKLEPKNGLPINFAIESWLTREAWRTTGINSDEQE-----VMIW-IYYD 205

DB 264 SDFGLSRGYSYRMNPHKKVPJR-----WLAPEVPR-TGFTPKDVFAYGVMCEVTHD 317
QY 206 GLQP-AGSKVKEIVVPIVNGTPVNNATFEVWKANIG---WEYVAFRIKTIKEGTVTI 259
DB 318 GIEPYFGMKVAB-VLPRVQNG--YRMPFE--ANVPPAIVRFITVTRICAGAEERVTM 369

RESULT 10

T21201

hypothetical protein F21D5.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T21201

R;Berks, M.

submitted to the EMBL Data Library, September 1995

A;Reference number: Z19389

A;Accession: T21201

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-713 <WIL>

A;Cross-references: EMBL:Z54271; PIDN:CAA91039.1; GSPDB:GN00022; CESP:F21D5.3

A;Experimental source: clone F21D5

C;Genetics:

A;Gene: CESP:F21D5.3

A;Map position: 4

A;Introns: 5/3; 41/3; 93/2; 180/1; 211/3; 252/1; 293/1; 332/3; 399/1; 459/3; 495/3; 628/3

Query Match 5.9%; Score 101; DB 2; Length 713;

Best Local Similarity 24.5%; Pred. No. 5.9;

Matches 70; Conservative 28; Mismatches 122; Indels 66; Gaps 15;

QY 14 ILLVQAIYFEVKYHTS---EDKSTGTSSTPQTTLSTTKVLKIRYPDGMWPGAP---- 66

DB 376 IPVYFGYIE-YETDGLPEDXTTFLMH-PECTQNKKTVLNCPFO---EMWAPNFTC 430

QY 67 --IDKDGDNPEFYIEINLNLNAT---GFAETYNLTSGVLHYVQQLDNIVLRDRSN 120

DB 431 ISYDKLENDPE-KTEA---EILQATQPNGGFEEHFIN-----MHDSQMDQ----- 473

QY 121 WHVHGYPEIFYGNKPNWANYATDGPPLPSKVSNLN-----DFYLTTSYKLEPKNG 170

DB 474 -----FLFPQPMGIPYHGDGMTETSTSCQATACANMSDLDTTCRCFYHLEHKLN 525

QY 171 LPINFAIESWLTREAWRT-----TCINSDEQVMI--WYYDGLQAPAGSKVKEIVVPIIV 223

DB 526 NLVQITLYNMLGGAMGTGYAHPFLHGHFFYVMKVGWPSYNGSGFIDQMNQIDCP--- 582

QY 224 NGTPYNATEVVKAN--IGWEYVAFRIKTIKEGTVTIPIYGAFISV 267

DB 583 -GRDVSCNGKRWKNDWLGGALEAWNTKNTKRTITLTPVGGYITI 627

RESULT 11

S12610

cellulase (EC 3.2.1.4) precursor - Aspergillus aculeatus

N;Alternate names: endo-1,4-beta-glucanase

C;Species: Aspergillus aculeatus

C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000

C;Accession: S12610; S14118; S40186; JQ0458

R;Ooi, T.; Shimmyo, A.; Okada, H.; Murao, S.; Kawaguchi, T.; Arai, M.

Nucleic Acids Res. 18, 5894, 1990

A;Title: Complete nucleotide sequence of a gene coding for Aspergillus aculeatus cellulase

A;Reference number: S12610; MUID:91016934; PMID:2216782

A;Accession: S12610

A;Molecule type: DNA

A;Residues: 1-237 <OOII>

A;Cross-references: EMBL:D00546; NID:g217818; PIDN:BA00435.1; PID:g217819

R;Ooi, T.; Shimmyo, A.; Okada, H.; Hara, S.; Ikenaka, T.; Murao, S.; Arai, M.

Curr. Genet. 18, 217-222, 1990

A;Title: Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase) from Aspergillus

A;Reference number: S14118; MUID:91064758; PMID:2249253

A;Accession: S14118

Qy 149 SKVSNL-----TDFYLTYSY-----KLEPNGLPI 173
Db 159 HDLYNMTFLLAQGLFOKGVTEALMAISYNGTIFGSEPTNVHRBELYFNQILKXNLP 218
Qy 174 --NPAIESWLTR-----AWRTTGINSDEQEWIWI-----YYDGLQPAGS--- 212
Db 219 PTNLSLLIYDQLEAHGINPWSGAEGGYEQHLHFAIFLSVAAYYGA---AGAAKL 275
Qy 213 -----KVKEIVPPIVNGTP-----VNATEVVKANIGWEVVAPRIKTIKEG 255
Db 276 SNELMYGLVNLNNVTQKIINETDNVFLQFYQSSVIPSPQSQSIWSALALVIK-----G 330
Qy 256 TVTIPYQAFISVAANIGSLNPNLYLELVEIGTFGTPTSTSAHLEWITNITLTPLDR 315
Db 331 QTVPEAG-----NWLAEYAAIWNTT-----TYPATQYPLNW--SNITLMANPF 373
Qy 316 P 316
Db 374 P 374

RESULT 13
CSMS
Complement C5 precursor - mouse
N;Contains: C5a anaphylatoxin; C5b
C;Species: Mus musculus (house mouse)
C;Date: 19-Nov-1988 #sequence revision 15-Oct-1994 #text_change 18-Jun-1999
C;Accession: A35530; A27538; A40429
R;Wetsel, R.A.; Fleischer, D.T.; Haviland, D.L.
J. Biol. Chem. 265, 2435-2440, 1990
A;Title: Deficiency of the murine fifth complement component (C5). A 2-base pair gene
A;Reference number: A35530; MUID: 90153853; PMID:2303408
A;Accession: A35530
A;Molecule type: mRNA
A;Residues: 1-215, 'L' <WET>
A;Cross-references: GB:M35526; GB:J05234; NID: g192302; PIDN: AAA37348.1; PID: g309123
R;Wetsel, R.A.; Ogata, R.T.; Tack, B.F.
Biochemistry 26, 737-743, 1987
A;Title: Primary structure of the fifth component of murine complement.
A;Reference number: A27538; MUID: 87185363; PMID:2436653
A;Accession: A27538
A;Molecule type: mRNA
A;Residues: 'PGL', 44-1680 <WET2>
R;Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Wetsel, R.A.
J. Biol. Chem. 266, 11818-11825, 1991
A;Title: Structure of the murine fifth complement component (C5) gene. A large, highly
nt component genes.
A;Reference number: A40429; MUID: 91268053; PMID:1711041
A;Accession: A40429
A;Molecule type: DNA
A;Residues: 1-15 <HAV>
A;Cross-references: GB:M64852
C;Comment: Complement C5 contains two disulfide-linked chains, formed by removal of four
(beta and alpha' chains).
C;Comment: Activation of C5 initiates the spontaneous assembly of the late complement C
is the foundation upon which the membrane attack complex is assembled.
C;Comment: C5a has potent spasmogenic and chemotactic activity.
C;Genetics:
A;Map position: 2
A;Intons: 22/3; 86/3; 140/3; 164/3; 195/2; 223/1; 253/2; 291/3; 334/1; 372/3; 434/3; 5
3; 1224/1; 1282/3; 1343/3; 1364/3; 1392/1; 1411/2; 1445/3; 1470/3; 1506/1; 1534/1; 1564
C;Superfamily: alpha-2-macroglobulin
C;Keywords: complement alternate pathway; complement pathway; cytolysis; glycoprotein;
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-674, 679-1679/Product: complement C5 #status predicted <MAT>
F;19-674, 756-1679/Product: C5b #status predicted <CSB>
F;19-674/Product: complement C5 and C5b beta chain #status predicted <C5BB>
F;679-1679/Product: complement C5 alpha chain #status predicted <C5A>
F;679-755/Product: C5a anaphylatoxin #status predicted <C5r>
F;756-1679/Product: C5b alpha' chain #status predicted <C5Ba>
F;567-814, 635-670, 702-728, 703-735, 715-736, 870-1531, 1105-1163, 1379-1509, 1408-1524-1
F;915, 1119, 1633/Binding site: carbohydrate (Asn) (covalent) #status predicted
C;Genetics:

Result No.	Query			ID	Description
	Score	Match	Length		
1	112.5	6.6	264	1	GUNS_ERWCA
2	103	6.1	421	1	DM3L_MOUSE
3	99.5	5.8	597	1	PEPX_BACAA
4	99	5.8	237	1	GUN_ASFAC
5	99	5.8	1680	1	COS_MOUSE
6	97.5	5.7	710	1	PPCE_MOUSE
7	95.5	5.6	2316	1	PTPZ_RAT
8	94.5	5.6	247	1	FLA1_THEVO
9	94.5	5.5	772	1	CIPB_CLOTM
10	93	5.5	608	1	YD93_METJA
11	93	5.5	785	1	VP3S_SCHPO
12	92	5.4	417	1	IF_MOUSE
13	91	5.3	1853	1	CIFA_CLOTM
14	90.5	5.3	1481	1	APU_THIEET
15	90	5.3	463	1	GUN_BACSP
16	89.5	5.3	710	1	PPCE_PIG
17	89.5	5.3	2352	1	MOKC_SCHPO
18	89	5.2	703	1	CDGT_BACS2
19	88	5.2	1183	1	CNA_STAAU
20	87.5	5.1	593	1	CSG_METFE
21	87.5	5.1	1848	1	CBFA_CLOCL
22	87	5.1	312	1	YC90_METJA
23	87	5.1	485	1	ENT_ETNCO
24	87	5.1	3672	1	LM2L_CABEL
25	86.5	5.1	710	1	PPCE_BOVIN
26	86.5	5.1	783	1	TRF2_THEVO
27	86	5.1	239	1	GUNA_ASPPA
28	86	5.1	879	1	MANB_CAPHI
29	86	5.1	1256	1	ATL_STAAU
30	86	5.1	1664	1	RPAT_YEAST
31	85.5	5.0	626	1	GPBA_HUMAN
32	85.5	5.0	666	1	NEP1_THEVO
33	85.5	5.0	839	1	YDD8_HAEIN

```

Db 87 HSKAYPSLVSG-WHTAGYTENSGLP--QLSSNKITSNTVYSIKATG--TYNAAVDI 141
QY 180 WL---TREAWTTGINSDEQWMTWYVDGLOPAGSKVKIIVPILVNGTVPVNAATEVVK 236
Db 142 WFHTTKANWD-----SSPTDEIMWLJNDTNAAGPDYIETVFLG-----DSSMNVFK 189
QY 237 ANI-----GWEYVAFRIKTPKEGTVTIPYGAISVAANISSLPNY-----TELY 281
Db 190 GWINADNGGWNVSF-VHT-----NSASLNIRHPTDYLVTQTKQWMSDEKY 237
QY 282 LEDVEIGTE-FG 292
Db 238 ISSVEFGTEIFG 249

RESULT 2
ID DM3L MOUSE
AC Q9CWR8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA (cytosine-5)-methyltransferase 3-like.
GN DNMT3L.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. PubMed=11306809;
RX MEDLINE=21203556; PubMed=11306809;
RA Apolla U., Lyle R., Krohn K., Antonarakis S.E., Peterson P.;
RT "Isolation and initial characterization of the mouse Dnmt3l gene.";
RL Cytogenet. Cell Genet. 92:122-126(2001).
[2]
RP SEQUENCE FROM N.A.
RA Shaoping X., Hata K., Li E.;
RT "Full-length cDNA of a murine Dnmt3-like gene.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stabili F., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustingich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[4]
RP FUNCTION
RX MEDLINE=21625316; PubMed=11719592;
RA Bourhis D., Xu G.L., Lin C.S., Bollman B., Bestor T.H.;
RT "Dnmt3l and the establishment of maternal genomic imprints.";
RL Science 294:2536-2539(2001).
CC -!- FUNCTION: Probably not catalytically active as it has lost the
CC active site residues. May function not directly as a DNA
CC methyltransferase but as a regulator of methylation at imprinted
CC loci. It is required specifically for the establishment of genomic

```

```

CC imprints but is dispensable for their propagation. It is essential
CC for the de novo methylation of single-copy DNA sequences.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in testis, thymus, ovary, and heart.
CC -!- SIMILARITY: Belongs to the C5-methyltransferase family.
CC -!- SIMILARITY: Contains 1 ADD-type zinc finger.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ404467; CAB94726.1; -
CC EMBL: AF220524; AAF73868.1; -
CC EMBL: AK010434; BAB26936.1; -
CC MGI: MGI:1859287; Dnmt3l.
CC GO: GO:0005737; Cytoplasm; IDA.
CC GO: GO:0005720; C:nuclear heterochromatin; IDA.
CC GO: GO:0005634; C:nucleus; ISS.
CC GO: GO:0019899; F:enzyme binding; ISS.
CC GO: GO:0016564; F:transcriptional repressor activity; ISS.
CC GO: GO:0006306; P:DNA methylation; ISS.
CC GO: GO:0001701; P:embryonic development (sensu Mammalia); IMP.
CC GO: GO:0006349; P:imprinting; ISS.
CC GO: GO:0007283; P:spermatogenesis; ISS.
CC KW Zinc-finger; Zinc; Metal-binding; Nuclear protein.
CC FT ZN FING 87 179 ADD-TYPE.
CC SQ SEQUENCE 421 AA; 47992 MW; 6C996D220C6F6D83 CRC64;
CC
CC Query Match 6.1%; Score 103; DB 1; Length 421;
CC Best Local Similarity 20.6%; Pred. No. 1.3;
CC Matches 32; Conservative 22; Mismatches 53; Indels 48; Gaps 5;
CC
CC QY 68 DXDGDGNPEFYIEINLW-----NIINATGFAEMTYNLTSGVLHYVQQLDN 112
CC Db 203 DQEGAGPEIYKTVSAWKQPVRLSLPRNIDKVLKSLGFLSGSGSGGGLTKYVEDVTN 262
CC QY 113 IVLRDRSNVHGYPEIFYGNKPNANYATDGPILPSKVSNLTDVLTISYKLPKNGLP 172
CC Db 263 VVRDVEKV--GFFDLVVGSTQ-----PLGSSCDRCPCGWINQFHRI----- 302
CC QY 173 INPAISWLTREAWRTTGINSDEQWMTWYVDGLOPAGSKVKIIVPILVNGTVPVNAATEVVK 207
CC Db 303 LQVALP-----RQSSQRPFFWIFMDNL 324
CC
CC RESULT 3
CC ID PEPX_BACAA STANDARD; PRT; 597 AA.
CC AC Q81PE9;
CC DT 10-OCT-2003 (Rel. 42, Created)
CC DT 10-OCT-2003 (Rel. 42, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Putative Xaa-Pro dipeptidyl-peptidase (EC 3.4.14.11) (X-Pro
CC dipeptidyl-peptidase) (X-prolyl-dipeptidyl aminopeptidase) (X-PDAP).
CC GN BA2860.
CC OS Bacillus anthracis (strain Ames).
CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CC OX NCBI_TaxID=198094;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=22608414; PubMed=12721629;
CC RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
CC RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
CC RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
CC RA Kolonay J.F., Seaman M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
CC RA DeBoy R.F., Madupu R., Daugherty S.C., Durkin A.S., Haft D.H.,
CC RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
CC RA Berton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
CC RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,

```


OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A. PubMed=2303408;
 RP MEDLINE=90153853; PubMed=2303408;
 RA Wetzel R.A., Fleischer D.T., Haviland D.L.;
 RT "Deficiency of the murine fifth complement component (C5). A 2-base
 pair gene deletion in a 5'-exon.";
 RL J. Biol. Chem. 265:2435-2440(1990).
 RN [2]
 RP SEQUENCE OF 41-1680 FROM N.A.
 RX MEDLINE=87185363; PubMed=2436653;
 RA Wetzel R.A., Ogata R.T., Tack B.F.;
 RT "Primary structure of the fifth component of murine complement.";
 RL Biochemistry 26:737-743(1987).
 CC -!- FUNCTION: Activation of C5 by a C5 convertase initiates the
 CC spontaneous assembly of the late complement components, C5-C9,
 CC into the membrane attack complex. C5b has a transient binding site
 CC for C6. The C5b-C6 complex is the foundation upon which the lytic
 CC complex is assembled.
 CC -!- FUNCTION: Derived from proteolytic degradation of complement C5,
 CC C5 anaphylatoxin is a mediator of local inflammatory process. It
 CC induces the contraction of smooth muscle, increases vascular
 CC permeability and causes histamine release from mast cells and
 CC basophilic leukocytes. C5a also stimulates the locomotion of
 CC polymorphonuclear leukocytes (chemotaxis) and direct their
 CC migration toward sites of inflammation (chemotaxis).
 CC -!- SUBUNIT: C5 precursor is first processed by the removal of 4 basic
 CC residues, forming two chains, beta and alpha, linked by a
 CC disulfide bond. C5 convertase activates C5 by cleaving the alpha
 CC chain, releasing C5a anaphylatoxin and generating C5b (beta chain
 CC + alpha chain).
 CC -!- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.
 CC -!- SIMILARITY: Contains 1 anaphylatoxin-like domain.
 CC -!- SIMILARITY: Contains 1 NTR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M35525; AAA37349.1; -;
 DR EMBL; M35526; AAA37348.1; -;
 DR PIR; A35530; C5MS.
 DR HSP; P01031; 1KJS.
 DR MGD; MGI:96031; Hc.
 DR InterPro; IPR002890; A2M N.
 DR InterPro; IPR009048; AM_receptor_bind.
 DR InterPro; IPR000020; Anaphylatoxin.
 DR InterPro; IPR001840; Anaphylatoxin.
 DR InterPro; IPR001599; MacroglobulinA2.
 DR InterPro; IPR001134; Netrin C.
 DR InterPro; IPR008930; Terp_cyc_toroid.
 DR Pfam; PF00207; A2M; 1.
 DR Pfam; PF01835; A2M N; 1.
 DR Pfam; PF01821; ANA7O; 1.
 DR Pfam; PF01759; NTR; 1.
 DR PRINTS; PRO0004; ANAPHYLATOXN.
 DR ProDom; PD003264; Anaphylatoxin; 1.
 DR SMART; SM00104; ANATO; 1.
 DR SMART; SM00643; C345C; 1.
 DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; FALSE_NEG.
 DR PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
 DR PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
 DR PROSITE; PS01889; NTR; 1.
 DR Complement pathway; Complement alternate pathway; Glycoprotein;
 KW Plasma; Membrane attack complex; Cytolysis; Inflammatory response;
 KW Signal.

FT SIGNAL 1 18
 FT CHAIN 19 1680
 FT CHAIN 19 674
 FT PROPEP 675 678
 FT CHAIN 679 1680
 FT PEPTIDE 679 755
 FT CHAIN 756 1680
 FT DOMAIN 702 736
 FT DOMAIN 1536 1679
 FT DISULFID 702 728
 FT DISULFID 703 735
 FT DISULFID 715 736
 FT CARBOHYD 427 427
 FT CARBOHYD 915 915
 FT CARBOHYD 1119 1119
 FT CARBOHYD 1633 1633
 FT VARIANT 216 216
 FT VARIANT 217 1680
 FT SEQUENCE 1680 AA; 188877 MW; 81E5A16FAC7D95C CRC64;
 SQ
 Query Match 5.8%; Score 99; DB 1; Length 1680;
 Best Local Similarity 22.8%; Pred. No. 15;
 Matches 66; Conservative 32; Mismatches 95; Indels 96; Gaps 13;
 QY 81 INLWNI-----LNATGFAEMTYNLTSGVLYHYVQOLDNIVLRDRSNWVHGYEIPYG--- 131
 DB 1 MGLWGLICLLFLDKTWGQEQYIVISAPKILRVGSSENVVIG-----VHGYTEAFDATLS 55
 QY 132 -----NKPNNVATDGPPLPSKYSNLTDFYL-TISYKLEPKN 169
 DB 56 LKSYDPKKVTFSSGVYNLSPENKFNQAAALLTLPNQVPRESVSHVLEVVVSKHFSKSK 115
 QY 170 GLPINFAIESWLTREAWRTGI-----NSDRQEVMIWY--DGLQFA----- 210
 DB 116 KIPITY-----NNGILFIHTDKPVYTPQSKIRVYSLGDDLKPAKRETVLTF 163
 QY 211 ----GSKVKEIVVPIVING-----TPVNATFEW--KANIG-----WEY 243
 DB 164 IDPEGSEV-DIVSENDYTGIIISFPDFKIPSPKYGWTKIKANYKKDFTTGTAYFEIKY 222
 QY 244 VAFRIKTPKEGTIVPYGAFISVAANTSSLPNTLYLEDVEIGTEFG 292
 DB 223 VLPFRSVSIELERTFIGNKFNKFNFTVKARFYKNV-VFDAEYVAFG 270
 RESULT 6
 PPCE MOUSE STANDARD; PRT; 710 AA.
 AC QOQR6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Prolyl endopeptidase (EC 3.4.21.26) (Post-proline cleaving enzyme)
 DE (PE).
 GN PREP OR PEP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98207037; PubMed=9538240;
 RX Ishino T., Ohtsuki S., Homma K., Natori S.;
 RT "cDNA cloning of mouse prolyl endopeptidase and its involvement in DNA
 RT synthesis by swiss 3T3 cells.";
 RL J. Biochem. 123:540-545(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99377037; PubMed=10446174;
 RA Kimura A., Yoshida I., Takagi N., Takahashi T.;
 RT "Structure and localization of the mouse prolyl oligopeptidase gene.";
 RL J. Biol. Chem. 274:24047-24053(1999).

[3] SEQUENCE FROM N.A.
RC STRAIN=FVE/N; TISSUE=Limb, and Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozny K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Cleaves peptide bonds on the C-terminal side of prolyl
CC residues within peptides that are up to approximately 30 amino
CC acids long.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of Pro-|-Xaa >> Ala-|-Xaa in
CC oligopeptides.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILILARY: Belongs to peptidase family S9A.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB007631; BAA8239.1; -
CC EMBL; AB022053; BAA83071.1; -
CC EMBL; AB022047; BAA83071.1; JOINED.
CC EMBL; AB022048; BAA83071.1; JOINED.
CC EMBL; AB022049; BAA83071.1; JOINED.
CC EMBL; AB022050; BAA83071.1; JOINED.
CC EMBL; AB022051; BAA83071.1; JOINED.
CC EMBL; AB022052; BAA83071.1; JOINED.
CC EMBL; BC012869; AAH12869.1; -
CC EMBL; BC050830; AAH50830.2; -
CC PIR; JWC080; JWC080.
CC HSP; P23687; 10FS.
CC MEROPS; S09.001; -
CC MGD; MGI:1270863; Prep.
CC GO; GO:0005737; Cytoplasm; IDA.
CC GO; GO:0005634; C:nucleus; IDA.
CC InterPro; IPR002471; Pept_S9_AS.
CC InterPro; IPR001375; Peptidase_S9.
CC InterPro; IPR002470; Peptidase_S9A.
CC InterPro; IPR004106; Peptidase_S9A_N.
CC InterPro; IPR000379; Ser-estr.
CC Pfam; PF00326; Peptidase_S9; 1.
CC Pfam; PF02897; Peptidase_S9_N; 1.
CC PRINTS; PR00862; PROLIGOPTASE.
CC PROSITE; PS00708; PRO-ENDOPEP_SER; 1.
KW Hydrolase; Serine protease.
FT ACT_SITE 254 554
FT ACT_SITE 641 641 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 680 680 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 710 AA; 80751 MW; 13010D5DCAV3C0E CRC64;

Query Match 5.7%; Score 97.5; DB 1; Length 710;
Best Local Similarity 19.9%; Pred. No. 6.5;

Matches 69; Conservative 52; Mismatches 130; Indels 95; Gaps 16;
QY 22 FVEKHTSDKSTNTSTSTPQTTL-----STTKVLKIRYPDDGEW-PCAPIDKQGD 72
DB 186 FYNFPOQDKSDGTSTETSNLHKLCHYVLGTQSDSDILCAEFDPDPKMWGGAELSDDO 245
QY 73 GNPEFYIENLW---NILNATGAEMTY--NUTSGVLHYVQOLDNIVLRDNRNWHGVYPE 127
DB 246 ----YVLLSIWEGCDPVNRLMYCLOQEPNGITGILKWKLIDN--FEGEYDYVTNEGT 298
QY 128 IFYGNKPWNA-NVA---TDGPILPSK-----VSNLTDPLYLTISKLEBPK 168
DB 299 VTFKTRNPNRLINDTDPDESKVLPVEHEKDVLEWVAVCURSNFLVLCYLHVK 358
QY 169 NGLPINFALESWLTR----EAWRTTGINSDEQEVMIWYDGLQAG-----SKVKEIVV 219
DB 359 NILQLHDLTGTGALLKTFPLDVGVSGRKKDSIFYQTSFLSPGVYIHCDLTKEELE 418
QY 220 PIIVNGTPVNATFEVWKANIGWEYVAPRIKTPKEGSTVIP-----260
DB 419 PMWFREVTY-----KGIDAADYQTIQIFYPSKQGT-KIPMFIVHKKGIKLDGSHAPL 470
QY 261 --YGAFISVAANISSLPNTY-----ELYLEDVEIGTEFG 292
DB 471 YGVGGF-----NISITPNYSVRLIFVRHMGGLAVANIRGGGEYV 511
RESULT 7
PTPZ_RAT
ID PTPZ_RAT STANDARD; PRT; 2316 AA.
AC Q62656; Q62621;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Receptor-type protein-tyrosine phosphatase zeta precursor
DE (BC 3.1.3.48) (R-PTP-zeta) (Phosphacan) (3F8 chondroitin sulfate
DE proteoglycan) (3H1 keratan sulfate proteoglycan).
DE PTPRZ1 OR PTPRZ OR PTPZ.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=96063026; PubMed=7579589;
RA Maurel P., Meyer-Puttlitz B., Flad M., Margolis R.U.,
RA Margolis R.K.;
RT "Nucleotide sequence and molecular variants of rat receptor-type
RT protein tyrosine phosphatase-zeta/beta.";
RL DNA Seq. 5:323-328(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3), AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=94195772; PubMed=7511813;
RA Maurel P., Rauch U., Flad M., Margolis R.K., Margolis R.U.;
RT "Phosphacan, a chondroitin sulfate proteoglycan of brain that
RT interacts with neurons and neural cell-adhesion molecules, is an
RT extracellular variant of a receptor-type protein tyrosine
RT phosphatase.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:2512-2516(1994).
RN [3]
RP BINDING TO N-CAM AND NG-CAM (PHOSPHACAN).
RX MEDLINE=95096181; PubMed=7528221;
RA Milev P., Friedlander D.R., Sakurai T., Karthikeyan L., Flad M.,
RA Margolis R.K., Grumet M., Margolis R.U.;
RT "Interactions of the chondroitin sulfate proteoglycan phosphacan, the
RT extracellular domain of a receptor-type protein tyrosine phosphatase,
RT with neurons, glia, and neural cell adhesion molecules.";
RL J. Cell Biol. 127:1703-1715(1994).
RN [4]
RP BINDING TO TENASCIN (PHOSPHACAN).
RX MEDLINE=94216329; PubMed=7512960;

Query Match 5.5%; Score 94; DB 1; Length 772;
Best Local Similarity 23.3%; Pred. No. 13;
Matches 66; Conservative 30; Mismatches 127; Indels

```

QY 34 TSNTSSTPQTTLTKYKIRYPDDGWPAGAPDKDGNPPEVIEINLWNLNATGPA 93
DB 265 TNPNTANTPVSNG--LKVEF-----YNENPSDTTNSINPQKV-----TNTG 307
QY 94 EMTYNLTSGVLHYVQOQDNIVLRDRSMVHGYPRIFYGNKPWNA-NYATDGPILPSPKVS 152
DB 308 SSAIDLSKLTLYYYTVDG--QKDOTFWC-DHAAIIGSGSYNGITSNVKGTFFVKMSST 364
QY 153 NLTDYLYTISY---KLEPKXGLPI--NEAIESLWLTREARWTTGINSDEQ----- 196
DB 365 NNADTYLEISFTGGTLEGAHVQIQGRFAKNDWSNYTSDNSYFSKRSQFVENDQVTAYL 424
QY 197 -EVMIIWYDGLQAGSKVEIIVPVIIVNGTPVNATFEVWKANIGWEYVAFRIKTIKEG 255
DB 425 NGVLWV---GKEPGGS-----VVPSTQVTPPTATTTPPSDDPNAIK--IKVD 473
QY 256 TWTIPYGFISVANISLIP-----NYTELYLEDVEI 287
DB 474 TVNAKPGDVTNIPVRFSGIPSKGIANCDFVSYDPNVLEIIEI 516

RESULT 10
YD93_METJA
ID YD93_METJA STANDARD; PRT; 608 AA.
AC Q58788;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1393.
GN MJ1393.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.P., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts C.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: TO M.JANNASCHII MJ1394 AND A.FULGIDUS AF2028.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67579; AAB99403.1; -.
CC F01; H64473; H64473.
CC TIGR; MJ1393; -.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
SQ SEQUENCE 608 AA; 56768 MW; 010FAFLC39F8C73C CRC64;
Query Match 5.5%; Score 93; DB 1; Length 608;
Best Local Similarity 20.1%; Pred. No. 12;
Matches 82; Conservative 54; Mismatches 128; Indels 144; Gaps 22;
QY 4 KKFIVSVIL-TILLVQAIY-----FVEKYHTSEDKSTNTSSTPQTTLSTTKV 51

```

```

DB 2 RKLIFMALLMSLLFIGTVEGYGNGPLVYAYEYKNI-----TGNTT----- 43
QY 52 LKIRYPDDGWPAGAPDKDGNPPEVIE-INLWNLNATGPAEMTYNLTSGVLHYVQOL 110
DB 44 -----GDLVSTIESITGYIVINNTG--TTINDT---LYDVWVA 78
QY 111 DNIVLRDRSMVHGYPRIFYGNKPWNA-NYATDGPILPSPKVS 138
DB 79 VNI-----SNNTITG-PEVYVNGTPKGVFTSESSAPAYTNLP-NANTYIHIPILPNNSYVII 131
QY 139 -YATDGP-----PLSKVSNLTFDYLTISYKLE--PKNGLPINFAIESWL 181
DB 132 KFAIDKSIITGVPLIINETYSYDTSKISERLSNWSVNLNRSVNSALPATDTPSVIMTKYL 191
QY 182 TR-----EAW-----RTTGINSDEQEVMIWI-----YYDGLQFAG-----SKYEIV 219
DB 192 SNDPNNGYSDTWNFLNITGAINEGSIITLWDGPFYLPFGYNDLSLTWTGVVINTKNTATITI 251
QY 220 PIIVNGTPVNATFEVWK---ANIGWEYVAFRIKTIKEGTVTIPYGFISVAN--ISSL 274
DB 252 NITGNNTYTNRTGTLMKYGFVIFPEFNGTKSGTKI-EGIYATGYGGVATKEGPFNLAS 310
QY 275 PNTELYLEDVEITGTEFTPTSTSAHLEWNIIN-----ITLPLDRPLI 318
DB 311 SGKYEIWESEANVSKASSYVFNLTHTVTIWA VNGSNPVLDPFNITLL 358

RESULT 11
VP35_SCHPO
ID VP35_SCHPO STANDARD; PRT; 785 AA.
AC O74552; P78830;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vacuolar protein sorting-associated protein vps35.
GN VPS35 OR SPC777.13.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown N., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders R., Squares S., Stevens K.,
RA Skelton J., Simmonds M., Squares R., Tacey S.V., Warren T., Whitehead S.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels B., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux C., Moore K., Hurst S.M.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.B., Paulsen I., Potashkin J.,
RA Sipakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 279-785 FROM N.A.

```

```

RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomyces pombe
RL cDNAs."
CC DNA Res. 4:363-369(1997).
CC -!- FUNCTION: May play a role in vesicular protein sorting, analogous
CC to the yeast retromer proteins (By similarity).
CC -!- SUBUNIT: Probably a component of a membrane-associated
CC multiprotein complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Associated with an intracellular membrane
CC (By similarity).
CC -!- SIMILARITY: Belongs to the Vps35 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL031532; CAA20717.1; -.
CC EMBL; D89178; BAAL3840.1; -.
CC PIR; T11719; T11719.
CC GeneDB SPombe; SPCC77.13; -.
CC InterPro; IPR005378; Vps35.
CC Pfam; PF03635; Vps35; 1.
CC Transport; Protein transport; Membrane.
CC CONFLICT 337 337; OBA4BLCFBF58D68 CRC64;
CC SEQUENCE 785 AA; 90608 MW; 0BA4BLCFBF58D68 CRC64;
CC -----
Query Match 5.5%; Score 93; DB 1; Length 785;
Best Local Similarity 20.7%; Pred. No. 16;
Matches 75; Conservative 56; Mismatches 117; Indels 114; Gaps 19;
QY 8 IVSILTLLVQAIFYVEKYHSEDKSTNTSTPQTTLSTTKV-LKIRYPDGGEWGP 66
Db 244 VTQLVWAMNLTDTYVQREYES-DSSNEDESEVTE-KLGDIKINEVQKDEQCEGDK 301
QY 67 IDKDGNGNPEFYIEINLWN-----ILNATGP-AEMTYNLTSGLVHYVQQLDNIVLRDSN 120
Db 302 VIP-----PEVAIQELVSHVHVIEQSRGLPLDCIVLSILSIL-----NPLR---- 345
QY 121 WWHGYPEIFYGNKPNAN-----YATDGPILPISKVSNLTD-----FYLT 160
Db 346 -----CYPKQYADRVFPQYNEHIINQPSLSALHERPLOSALCAILLPLATYFPS 397
QY 161 ISYKLEPKNGPLPINFATESWLTREAWR-----
Db 398 PSYCLEQNPLFPVNAQDPNLRDIARNVQKITEKGHSISELTAQELLGFVSVITEKK 457
QY 190 GINS--DEQYMIWIIY-----DGLQAGSKYKEIVVPIIVNGTPVNA 230
Db 458 GVDSDLDLQNALMVHLYNDDPOIQIEILRSLKDTFKAGENVK-YLLFPVWNRCIFLA 516
QY 231 -TEFV-----W--KANTGWEVAVRIKTPKEG-TVPIPYGAFISVAANISLPLNYTELY 281
Db 517 RNFRIKCMDAEKVRLLWEFVNTCINLVYKNGSDLSLCLALYLS-AAEMADQENYDPFA 575
QY 282 LE 283
Db 576 YE 577

```

RESULT 12

IF MOUSE

ID IF MOUSE

AC IF MOUSE

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Intrinsic factor precursor (IF) (INF) (Gastric intrinsic factor).

```

GN GIF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=BALB/c; TISSUE=Stomach;
RX MEDLINE=94075347; PubMed=8253786;
RA Lorenz R.G., Gordon J.I.;
RT "Use of transgenic mice to study regulation of gene expression in the
RT parietal cell lineage of gastric units."
RL J. Biol. Chem. 268:26559-26570(1993).
CC -!- FUNCTION: Promotes absorption of the essential vitamin cobalamin
CC (Cbl) in the ileum by specific receptor-mediated endocytosis.
CC -!- TISSUE SPECIFICITY: Gastric mucosa.
CC -!- SIMILARITY: Belongs to the eukaryotic cobalamin transport proteins
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L24191; AAA37882.1; -.
CC EMBL; L24192; AAA37881.1; -.
CC PIR; A49684; A49684.
CC MGD; MGI:1202394; Gif.
CC InterPro; IPR002157; Cobalamin_bind.
CC Pfam; PF01122; Cobalamin_bind; 1.
CC PROSITE; PS00468; COBALAMIN_BINDING; 1.
CC Transport; Cobalt transport; Glycoprotein; Signal.
CC SIGNAL 1 18
CC CHAIN 19 417 INTRINSIC FACTOR.
CC DISULFID 26 246 BY SIMILARITY.
CC DISULFID 103 288 BY SIMILARITY.
CC DISULFID 143 182 BY SIMILARITY.
CC CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 311 311 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 417 AA; 45395 MW; 421713BAE6358DC2 CRC64;
CC -----
Query Match 5.4%; Score 92; DB 1; Length 417;
Best Local Similarity 21.8%; Pred. No. 8.8;
Matches 78; Conservative 41; Mismatches 133; Indels 106; Gaps 19;
QY 12 LAILLVQAIFYVEKYHSEDKSTNTSTPQTTLSTTKVLRYPDGEWGPAPIDK 69
Db 4 LTYLLSLVWAVAGTSTRAQSS-----CSVPPQQPVDGLQALMENSVTDSDFP----- 53
QY 70 DGDGNPEFYIEINLWNILNATGAEMY-----NLTSGLVHY-VQQLD----- 111
Db 54 ----NPSILIANLAGAVNEAQLTYQLMASDSANLTSGQLALTVALTSSCRDPGSK 109
QY 112 -NIVLDRSNWVHGP-----EIPYG-----NKPWANYATD---GPIPLP 148
Db 110 VSTLLKKMENWSPSPGAESSAFYGFGLAILALCOCKSSATLPIAVAFKTLMBFSPFN 169
QY 149 SKVSNLTDFYLTISYKLEPKNGPLPINFATESWLTREAWR-----TGINSDEQVMIWIIYD 205
Db 170 VDTGAVATLALTCMY-----NKIPVG-----SQENYEDLFGQALKAVEIKSLRIKAD 217
QY 206 G-----LQPAGSKYKEIVVPIIVNGTPVNAFVWKNIGWEYVAPRIKTPKEGTVTIPY 261
Db 218 GIIGDIYSTGLAMQALSV-----TPEQPT-KKWDG-----EKTMTILNEIKQGFQND- 265
QY 262 GAFISVAANISLPLNYTELYLEDVIG-----TEGTPSTTSIAHLEWITNLT 310
Db 266 ----MSTAIQLPLSKGKTYLDVPOVTCGPDHVEFPLTIDYTPVTPIS-----VSNITV 314

```



```
FT STRAND 467 468
FT TURN 470 471
FT STRAND 473 482
FT TURN 483 484
FT STRAND 488 489
FT TURN 491 492
FT TURN 494 495
FT STRAND 498 498
FT STRAND 503 504
FT STRAND 509 512
FT TURN 513 514
FT STRAND 515 518
FT STRAND 1220 1224
FT STRAND 1226 1229
FT TURN 1231 1232
FT STRAND 1234 1242
FT TURN 1246 1247
FT STRAND 1249 1249
FT STRAND 1251 1257
FT TURN 1260 1262
FT STRAND 1263 1270
FT TURN 1272 1273
FT STRAND 1279 1282
FT TURN 1283 1288
FT STRAND 1293 1299
FT TURN 1301 1302
FT STRAND 1306 1307
FT STRAND 1309 1309
FT STRAND 1313 1322
FT TURN 1324 1325
FT STRAND 1329 1342
FT TURN 1344 1345
FT STRAND 1348 1348
FT STRAND 1351 1354
FT STRAND 1356 1360
SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 5.3%; Score 91; DB 1; Length 1853;
Best Local Similarity 20.7%; Pred. No. 69;
Matches 62; Conservative 31; Mismatches 119; Indels 88; Gaps 14;

QY 34 TSNTSTPTQTLSTTKVLKRYPDGEGWPGAPDKDGDGNPEYIEINLWNILNATGFA 93
Db 353 TINTNPANTPVSGN--LKVEF-----YNSNPSETTNSINPKV-----TNTG 395
QY 94 EMTNLTGVLHYVQQLDNVLRDRSNVHGYPIFYGNKPWNA-NYATDGPPLPSKVS 152
Db 396 SSAIDLKSLTRYVYVTDG--OKDQTEWC-DHAAIIGSNGSYNGITSNVKGTFFVMSST 452
QY 153 NLTDPLYLTISY--KLBPKNGLPI--NFAIESWLTREAWRTTGINSDEQ----- 196
Db 453 NNAUTYLEISTGTLBPAGHVQIQGRFAKNDWENYQSDNYSFKSASQFVNDQVTAYL 512
QY 197 -EVMIIWYDGLQFAGSKV---KEIVVPPIVNGTFVNATFEVVKANIGWEYVAFRIKTP- 251
Db 513 NGVLAV---GKEFGSVVPSTQVPTTPATTKPPATT-----KPPA 550
QY 252 -----IKEGTVTPYGAFTSVAANISLSP-----NTELYLEIVEI 287
Db 551 TTIPSDPDNAIKIVDTNNAKPGDTVNPVRFSGIPSKGIANCDFVYSYDPNVLEIIEI 610

RESULT 14
APU THEET
ID - APU THEET STANDARD; PRT; 1481 AA.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Amylopullulanase precursor (Alpha-amylase/pullulanase) [Includes:
DE Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase);
DE Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase)]

DE (Alpha-dextrin endo-1,6-alpha-glucosidase)].
DE APU.
OS Thermoanaerobacter ethanolicus (Clostridium thermohydrosulfuricum).
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=1757;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=39E / ATCC 33223;
RX MEDLINE=90147689; PubMed=2302196;
RA Mathupala S.P., Saha B.C., Zeikus J.G.;
RT "Substrate competition and specificity at the active site of
RL amylopullulanase from Clostridium thermohydrosulfuricum.";
CC Biochem. Biophys. Res. Commun. 166:126-132(1990).
CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
CC linkages in pullulan and in amylopectin and glycogen, and the
CC alpha- and beta-limit dextrins of amylopectin and glycogen.
CC -|- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -|- SUBUNIT: Monomer.
CC -|- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
CC -|- SIMILARITY: Contains 2 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; M97665; AAA23201.1; -.
CC PIR; S28669; S28669.
CC HSSP; Q08751; 1BVZ.
CC InterPro; IPR006589; Alp_amy1_cat_sub.
CC InterPro; IPR006048; Alpha_amy1_C-
CC InterPro; IPR006047; Alpha_amy1_cat.
CC InterPro; IPR001480; B_lectin.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR004185; Glyco_hydro_13IG.
CC Pfam; PF02806; alpha-amy1ase_1.
CC Pfam; PF02903; alpha-amy1ase_C_1.
CC Pfam; PF00041; fn3_2.
CC SMART; SM00642; Amy1_1.
CC SMART; SM00632; Amy1_C_1.
CC SMART; SM00060; FN3_2.
CC Carbohydrate metabolism; Multifunctional enzyme; Hydrolase;
KW Glycosidase; Calcium-binding; Repeat; Signal.
FT SIGNAL 1 31
FT CHAIN 32 1481
FT DOMAIN 926 1015
FT DOMAIN 1158 1250
FT ACT_SITE 628 628
FT ACT_SITE 657 657
FT ACT_SITE 734 734
FT METAL 398 398
FT METAL 400 400
FT METAL 403 403
FT METAL 404 404
FT METAL 449 449
FT METAL 451 451
FT SEQUENCE 1481 AA; 166362 MW; CB2635960539CAID CRC64;

Query Match 5.3%; Score 90.5; DB 1; Length 1481;
Best Local Similarity 19.5%; Pred. No. 57;
Matches 52; Conservative 45; Mismatches 91; Indels 79; Gaps 15;

QY 62 WPGAPDK-----DGDGNPEFYIEINLWNILNATGPAETYNLTSGVLHVQ----- 108
```

```
Db 95 MEGGVQSQNLSLHSDSVVFFYNNYNTSSVTDSTKYTPPEKUPRIVGTQSAIGA 154
QY 109 -----QLDNIVLR-RSNWVHG-----PEIFYGNK-----PMNANYATDGP-----IPL 147
Db 155 GDDWKPETSTIMRDYKFNWVEYVYANVPKRYEFTVLTGSPWDINYLNGEQGNIP 214
QY 148 -PSKVSNTLDFYLTISYKL-----EPKNG-----P-INFAIESWLTR- 184
Db 215 NVAYDTKITFYDYSNHNWIDYNPLPGPDNNIYDLDKHTDHPFRFAFGAIGTGT 274
QY 185 -AWRTTGINSBOEWMIVYDGLQPGAGSKVKEIIVPIVNGTVPNATFEVWKANIGWEY 243
Db 275 VTLRQAKNHLES AKI-SYWDIDIKKTRTE-----VPMYKIGQSPDGOYEWYVKLSFDY 328
QY 244 VAFRIKTP-----IKEGVTTIPYG 262
Db 329 -----PTRIYYFILKDGTKTAYYG 348

RESULT 15
GUN_BACSP STANDARD; PRT; 463 AA.
AC P29019;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (Endo-K).
OS Bacillus sp. (strain KSM-330).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92121880; PubMed=1770347;
RA Ozaki K., Sumitomo N., Ito S.;
RT "Molecular cloning and nucleotide sequence of the gene encoding an
RT endo-1,4-beta-glucanase from Bacillus sp. KSM-330."
RL J. Gen. Microbiol. 137:2295-2305(1991).
RN [2]
RP SEQUENCE OF 56-75, AND CHARACTERIZATION.
RX MEDLINE=91259037; PubMed=2045781;
RA Ozaki K., Ito S.;
RT "Purification and properties of an acid endo-1,4-beta-glucanase from
RT Bacillus sp. KSM-330."
RL J. Gen. Microbiol. 137:41-48(1991).
CC -!- FUNCTION: THIS ACID ENDOGLUCANASE IS ACTIVE OVER AN EXTREMELY
CC NARROW RANGE OF PH VALUES, BETWEEN 4.5 AND 6.5, WITH AN OPTIMUM
CC PH AT 5.2.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- PTM: THE N- AND THE C-TERMINUS MAY BE SUBJECTED TO PROTEOLYSIS.
CC -!- MISCELLANEOUS: ONE TRP RESIDUE HAS BEEN PROVED TO BE INVOLVED IN
CC THE MECHANISM OF ACTION OF ENDO-K.
CC -!- SIMILARITY: Belongs to cellulase family D (family 8 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M68872; AAA22409.1; -
DR PIR; A44808; A44808.
DR InterPro; IPR002037; Glyco_hydro.8.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF01270; Glyco_hydro-8; 1.
DR PRINIS; PR00735; GLYDRLASE8.
DR PROSITE; PS00812; GLYCOSYL_HYDROL_F8; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL
1 27
POTENTIAL.
```

```
FT PROPEP 28 55 POTENTIAL.
FT CHAIN 56 463 ENDOGLUCANASE.
FT ACT_SITE 130 130 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 191 191 NUCLEOPHILE (POTENTIAL).
SQ SEQUENCE 463 AA; 51882 MW; 407FA54F5236C59E CRC64;

Query Match 5.3%; Score 90; DB 1; Length 463;
Best Local Similarity 19.3%; Pred. No. 14;
Matches 46; Conservative 34; Mismatches 58; Indels 100; Gaps 12;

QY 91 GFAEMTYNL-----TSGVLHYVQQLDNIVLR-----DR----- 118
Db 192 GDLDIAYSLLAHKQWSSGKINYLKEAQNMTTKGKASNVTKNGLNGDWGDKSTEDT 251
QY 119 ---SNVHGYPEIFY---GNKPNANYATDGPPLPSKVSNTLDFYLTISYKLEPKNGL-- 171
Db 252 RPSDMMMSHLRAFYEFTGDKTW-----LNVIDNLNTYTNFTNKYSPTGLIS 299
QY 172 -----PINTAIESWLTRAWRTTG-----INSDEQEVMIW 201
Db 300 DFVVKNPPOPAPKDFLDESKYTDSYYNASRVPLRVMDYAMYGEKRGKVISD--KVATW 357
QY 202 I-----YDGLQPGAGSKVKE-----IVPIIVNGTVPNATFEVWKANIGWEYV 244
Db 358 IKSKTGKNPSKIVDGYKLDGTNIGDYPTAVYVSPFIAAGT-TNSKNQEW-VNSGWDWM 413

Search completed: June 29, 2004, 20:12:05
Job time : 20 secs
```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2004, 20:06:49 ; Search time 46 Seconds
(without alignments)
2188.050 Million cell updates/sec

Title: US-09-914-543-46

Perfect score: 1701

Sequence: 1 MSKKFVIVSILFILLVQAI.....HLEWITNTLPLDRPLIS 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_rvirus: *
16: sp_bacteriaph: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1701	100.0	319	17 Q9V2T0	Q9V2T0 pyrococcus
2	472	27.7	274	2 P66492	P66492 thermotoga
3	464	27.3	274	16 Q60033	Q60033 thermotoga
4	463.5	27.2	257	2 Q60032	Q60032 thermotoga
5	463.5	27.2	258	16 Q9S5X8	Q9S5X8 thermotoga
6	439.5	25.8	334	17 Q97X08	Q97X08 sulfolobus
7	430	25.3	332	17 Q97VG7	Q97VG7 sulfolobus
8	415.5	24.5	257	2 P96491	P96491 thermotoga
9	415.5	24.4	257	2 Q98428	Q98428 thermotoga
10	221.5	13.0	322	17 Q97V57	Q97V57 sulfolobus
11	143.5	8.4	261	2 Q7X494	Q7X494 bacillus 11
12	128	7.5	260	2 Q33897	Q33897 rhodothermu
13	114.5	6.7	381	2 Q54331	Q54331 streptomyce
14	111.5	6.6	269	16 Q82K36	Q82K36 streptomyce
15	111	6.5	239	3 Q74705	Q74705 aspergillus
16	111	6.5	1288	17 Q96XM6	Q96XM6 sulfolobus

17	110.5	6.5	264	2	O31030
18	109	6.4	304	16	Q9EP57
19	108	6.3	371	2	Q9KIH1
20	107.5	6.3	381	16	Q9RJY3
21	107	6.3	1612	11	Q9WUT8
22	104.5	6.1	1062	17	Q96X97
23	103.5	6.1	382	2	Q59963
24	103	6.1	434	16	Q7WLF6
25	103	6.1	859	5	Q21401
26	102.5	6.0	533	5	Q18684
27	101	5.9	713	5	Q19687
28	100.5	5.9	534	12	Q67668
29	100.5	5.9	587	12	Q67667
30	100.5	5.9	732	12	Q67666
31	100	5.9	263	16	Q9AN41
32	100	5.9	597	17	Q97IQ1
33	99.5	5.8	1041	16	Q8EU25
34	99.5	5.8	1043	16	Q8EV01
35	99	5.8	320	10	Q9ATK2
36	99	5.8	384	2	Q8X602
37	99	5.8	620	17	Q97UF5
38	99	5.8	660	5	Q86K14
39	99	5.8	742	11	Q8BNV3
40	99	5.8	2029	17	Q8TI44
41	98.5	5.8	390	5	Q86JF6
42	98.5	5.8	561	16	Q8NY14
43	98.5	5.8	572	16	Q9X20
44	97.5	5.7	279	10	Q49899
45	97.5	5.7	377	2	O08468

ALIGNMENTS

RESULT 1

Q9V2T0 PRELIMINARY; PRT; 319 AA.
ID Q9V2T0
AC Q9V2T0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endoglucanase A precursor (Endo-1,4-beta-glucanase b).
GN EGLA OR PF0854.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99084968; PubMed=9864341;
RA Bauer M.W., Driskill L.E., Callen W., Snead M.A., Mathur B.J.,
RA Kelly R.M.;
RT "An endoglucanase, EglA, from the hyperthermophilic archaeon
RT Pyrococcus furiosus hydrolyzes beta-1,4 bonds in mixed-linkage
RT (1-->3), (1-->4)-beta-D-glucans and cellulose.";
RL J. Bacteriol. 181:284-290(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb P.T., Brown J.R.;
RA "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RL ENBL; AF181032; AAD54602.1; -
DR ENBL; AE010200; AAL80978.1; -
DR GO; GO:000810; F:cellulase activity; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008985; ConA like lec gl.
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
DR Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 319 ENDOGLUCANASE A.

O31030 pectobacter
Q9EP57 oceanobacil
Q9KIH1 streptomyce
Q9RJY3 streptomyce
Q9WUT8 mus musculu
Q96X97 sulfolobus
Q59963 streptomyce
Q7WLF6 bordetella
Q21401 caenorhabdi
Q18684 caenorhabdi
Q19687 caenorhabdi
Q67668 goose parvo
Q67667 goose parvo
Q67666 goose parvo
Q9AN41 bradyrhizob
Q97IQ1 sulfolobus
Q8EU25 mycoplasma
Q8EV01 mycoplasma
Q9ATK2 arabidopsis
Q8X602 streptomyce
Q97UF5 sulfolobus
Q86K14 dictyosteli
Q8BNV3 mus musculu
Q8TI44 methanosarc
Q86JF6 dictyosteli
Q8NY14 staphylococ
Q9X20 staphylococ
Q49899 medicago sa
O08468 streptomyce

SQ SEQUENCE 319 AA; 35988 MW; 51686BCA569F7A43 CRC64;
Query Match 100.0%; Score 1701; DB 17; Length 319;
Best Local Similarity 100.0%; Pred. No. 3.3e-128;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSKKFVIVSILTIILLVQAIYVEKYHTSEDKSTNTSTPTQTLSTTKYVKIRYDDG 60
DB 1 MSKKFVIVSILTIILLVQAIYVEKYHTSEDKSTNTSTPTQTLSTTKYVKIRYDDG 60
QY 61 EWPGAPIDKGDGNPEFEIENLNLNATGFAEMTYNLTSGVLHYVQOLNIVLRDSN 120
DB 61 EWPGAPIDKGDGNPEFEIENLNLNATGFAEMTYNLTSGVLHYVQOLNIVLRDSN 120
QY 121 VWHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180
DB 121 VWHGYPEIFYGNKPNWANYATDGP:PLPSKVSNTLDFYLTISYKLEPKNGLPINFAIESW 180
QY 181 LTREAWRTGINSDEQVMIWYDGLQPGAGSKYKEIWPVIIVNGTPVNAATFEVWKANIG 240
DB 181 LTREAWRTGINSDEQVMIWYDGLQPGAGSKYKEIWPVIIVNGTPVNAATFEVWKANIG 240
QY 241 WEYVAFRIKTPKEGTVPYPGAFISVAANISSLPNTYLYLEDVEIGTEFGTPTTSAH 300
DB 241 WEYVAFRIKTPKEGTVPYPGAFISVAANISSLPNTYLYLEDVEIGTEFGTPTTSAH 300
QY 301 LEWITNITLPLDRPLIS 319
DB 301 LEWITNITLPLDRPLIS 319

RESULT 2

P36492
ID P36492 PRELIMINARY; PRT; 274 AA.
AC P36492;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase B (EC 3.2.1.4) (Cellulase)
DE (Carboxymethyl cellulase).
GN CELB.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2337;
RN [1]
RP SEQUENCE FROM N.A.
RA Bok J.D., Eveleigh D.E.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-137 FROM N.A.
RC STRAIN=Z2706-MC24;
RA Zverlov V.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-137 FROM N.A.
RC STRAIN=Z2706-MC24;
RX MEDLINE=93356813; PubMed=8352795;
RA Dakhova O., Kurepina N., Zverlov V., Svetlichnyi V.,
RA Velikodvorskaya G.;
RT "Cloning and expression in Escherichia coli of Thermotoga neapolitana
genes coding for enzymes of carbohydrate substrate degradation";
RL Biochem. Biophys. Res. Commun. 194:1359-1364(1993).
CC -! CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
DR EMBL; U93354; AAC95060.1; -;
DR EMBL; Z86103; CAB06782.1; -;
DR GO; GO:0008810; P:cellulase activity; IEA.
DR GO; GO:0016798; P:hydrolyase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR000985; ConA like lec.gl.
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF01670; Glyco_hydro_12; 1.

DR ProDom; PD004316; Glyco_hydro_12; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 274 AA; 31752 MW; 6B21C4F8CA7C9AA2 CRC64;
Query Match 27.7%; Score 472; DB 2; Length 274;
Best Local Similarity 40.4%; Pred. No. 8.4e-30;
Matches 93; Conservative 47; Mismatches 80; Indels 10; Gaps 4;
QY 79 IEINLWNLNATGFAEMTYNLTSGVLHYVQOLNIVLRDSNHWGYPYFVYGNKPNW 138
DB 39 MELFNWVKSVEG-ETWLKFGKQVQFYADIYVNIQNPDSVWVGYPYFVYGNKPNW 96
QY 139 YATDGPILPSKVSNTLDFYLTISYKLEPKNGLPINFAIESWLTREAWRTGINSDEQV 198
DB 97 --NSGTEILFVKVQKLPDFYVTLDSIYVNDLP:NLAMETWITKPKQTS-VSSGDVEI 153
QY 199 MIWYVYDGLQPGAGSKYKEIWPVIIVNGTPVNAATFEVWKANIGWEYVAFRIKTPKEGT 258
DB 154 MWYFNILNLPFGQKVDFTTIEINGSVETKMDVYPAPWGDYLAFLTTPMDGKGVK 213
QY 259 IPYGAFFISVAANI-----SSLPNTYLYLEDVEIGTEFGTPTTSAHLEW 303
DB 214 FNVKDFEKAABEVKIKGSTRVENFDEMYFCVWEIGTEFGDPNTTAAKFGW 263
RESULT 3
Q60033 PRELIMINARY; PRT; 274 AA.
AC Q60033;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase precursor (EC 3.2.1.4) (Cellulase)
DE (Endoglucanase) [Carboxymethyl cellulase].
GN CELB OR TMI525.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8;
RX MEDLINE=96425879; PubMed=8828221;
RA Liebl W., Ruille P., Bronnenmeier K., Riedel K., Lottespeich F.,
RA Greif I.;
RT "Analysis of a Thermotoga maritima DNA fragment encoding two similar
thermostable cellulases, Cella and CelB, and characterization of the
recombinant enzymes";
RL Microbiology 142:2532-2542(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX STRAIN=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima";
RL Nature 399:323-329(1999).
DR EMBL; Z69341; CAA93274.1; -;
DR EMBL; AF001800; AAD36592.1; -;
DR FIR; A72241; A72241.
DR TIGR; TMI525; -;
DR GO; GO:0008810; P:cellulase activity; IEA.
DR GO; GO:0016798; P:hydrolyase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008985; ConA like lec.gl.
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.


```

RESULT 6
Q97X08      PRELIMINARY;          PRT;      334 AA.
AC  Q97X08;
DT  01-OCT-2001 (TEMBLrel. 18, Created)
DT  01-OCT-2001 (TEMBLrel. 18, Last sequence update)
DT  01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE  Endoglucanase precursor (EC 3.2.1.4).
GN  SS01949.
OS  Sulfolobus solfataricus.
OC  Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC  Sulfolobus.
OX  NCBI_TaxID=2287;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 35092 / DSM 1617 / P2;
RX  MEDLINE=21332296; PubMed=11427726;
RA  She C., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA  Aweyer M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA  De Moors A., Erauso G., Flecher C., Gordon P.M.K.,
RA  Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA  Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA  Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA  Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT  "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL  Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR  EMBL; AB006803; AAK42142.1; -.
DR  PIR; G90360; G90360.
DR  GO; GO:000810; F:cellulase activity; IEA.
DR  GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR  GO; GO:0005275; P:carbohydrate metabolism; IEA.
DR  GO; GO:0005272; P:polysaccharide catabolism; IEA.
DR  InterPro; IPR008985; ConA like lec.gl.
DR  InterPro; IPR002594; Glyco_hydro_12.
DR  Pfam; PF01670; Glyco_hydro_12; 1.
DR  ProDom; PD004316; Glyco_hydro_12; 1.
DR  KEGG; Glycosidase; Complete proteome.
KW  Hydrolyase.
SQ  SEQUENCE 334 AA; 37619 MW; CB3F61393639A48 CRC64;

      Query Match      25.8%; Score 439.5; DB 17; Length 334;
      Best Local Similarity 33.4%; Pred. No. 4.e-27;
      Matches 117; Conservative 57; Mismatches 121; Indels 55; Gaps 17

Qy  1  MSKKKPVIVSIITLLV----QAIYFVEKYHTSDKSTN-----TSSTP 41
Db  3  MNKLYIIIVPIIVVGVIGAIYL---HHQSPNVAKTSITVTWNETLLMSITITVP 59
Qy  42  PQTITLSTTKVLRIPDDGEPGAPIDKDGCPPEFYIEINLAINLNGFRAEMYNLTS 101
Db  60  TTVTPTTSSIPGLIIVTSSASPTFVILNNTSVESFYLEVMNNAKTNGNTWTFNPLT 119
Qy  102  GVLIHYVQQDNLVLDRSNWVHGYPFIYGNKPNANYATDGPILPFSKVSNLT----DF 157
Db  120  RTLVSFNLTVQV--NPLQWINGPFIYVVGKPNDSYA--GNI-FPMIGNWTFFWVSF 173
Qy  158  YLTISYKLPKNGLPINF--AIESWLTFEAWR---TTGINSDEQVWIIYDGIQPGAS 212
Db  174  YINLT-KLDPS----INFDIASDAWIRPQIATSPGTAPGNGDIEIWWLFSQNLQPAQG 228
Qy  213  KVKIEIVPIIVNGTVPNATFEVK-ANI---GNEYAVAFRIKTP----IKEGVTI-PYGAF 264
Db  229  QVGEVVIPIYINHTLVNATTFQWKNKVPVGGHEIYAFR---PDGKWVINGVAVEPNLF 285
Qy  265  ISVAANISSLPNYLEYLDEVEIGTFTGT-PSTTSAHLEWMTITLTPL 313
Db  286  IKALANFASY-NITNYLTDEWFGTNGTMTSGTAYFSWTISNFYEITLL 334

RESULT 7
Q97YG7
AC  Q97YG7;
ID  Q97YG7;

```

```
OS Thermotoga neapolitana.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
CX NCBI_TaxID=2337;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2706-MC24;
RA Zverlov V.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2706-MC24;
RX MEDLINE=93356813; PubMed=8352795;
RA Dakhova O., Kurepina N., Zverlov V., Svetlichnyi V.,
RA Velikodvorskaya G.;
RT "Cloning and expression in Escherichia coli of Thermotoga neapolitana
RT genes coding for enzymes of carbohydrate substrate degradation.";
RL Biochem. Biophys. Res. Commun. 194:1359-1364(1993).
DR EMBL; Z86103; CAB06783.1; -.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008985; ConA_like_1ec.gi.
DR Pfam; PF01670; Glyco_hydro_12.
DR ProDom; PD004316; Glyco_hydro_12; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 257 AA; 30192 MW; D6E148CABAC864F7 CRC64;

Query Match 24.5%; Score 416.5; DB 2; Length 257;
Best Local Similarity 33.6%; Pred. No. 2.2e-25;
Matches 86; Conservative 51; Mismatches 100; Indels 19; Gaps 6;

QY 59 DGEWPGPIDKDGNGNPEFYIEINLWLNATGTAETYNLTSGVLHYVQQLDNIVLRDR 118
DB 11 DFRNDMPLS-----MELNLNMYRYTGTVMRFD--GERLTFNGDVEDLSAREP 58

QY 119 SNWVHGYPEIFYGNKPNANYATDGPILPSPKVSNLTDFTYLTISYKLEPKNGLPINFAIE 178
DB 59 ERYILGYPEFYGYKPEW-RHAAEG-TKLPLVSVSVESFTVELSFEIDHMPSLPLNFAME 116

QY 179 SWLTREAWRTTGINSDEQEWMIWYDGLQAGSKVKEIVPIVNGTPVYNATFEVWKAN 238
DB 117 TWLTREKYQVEASIGD-VEIMVWFYFNELTGPGKKGVEYTVSFELNGEKGKGIWELWHA 175

QY 239 IGWEYVAFRIKTIKEGTVPYGFATISVAANISS-----LPNYTELYLEDVEIGTEFGTP 294
DB 176 WNWDFLAFRLKPNVRKGRVRFNVKDFLDVAGEYLSRSTRVKDFDLYFTVWEIGTEFGSP 235

QY 295 STTSAHLEWITNITL 310
DB 236 ETKSARFGWTFNFSI 251

RESULT 9
OC08428 PRELIMINARY; PRT; 257 AA.
AC O08428;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase A (EC 3.2.1.4).
GN CELA.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
CX NCBI_TaxID=2337;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054912; PubMed=9835561;
RA Bok J.D., Yernool D.A., Eveleigh D.E.;
RT "Purification, characterization, and molecular analysis of
RT thermostable cellulases CELA and CelB from thermotoga neapolitana.";
RL Appl. Environ. Microbiol. 64:4774-4781(1998).
```

```
DR EMBL; U93354; AAC95059.1; -.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008985; ConA_like_1ec.gi.
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 257 AA; 30201 MW; 14856B9EBAD0EDEF CRC64;

Query Match 24.4%; Score 415.5; DB 2; Length 257;
Best Local Similarity 33.6%; Pred. No. 2.6e-25;
Matches 86; Conservative 51; Mismatches 100; Indels 19; Gaps 6;

QY 59 DGEWPGPIDKDGNGNPEFYIEINLWLNATGTAETYNLTSGVLHYVQQLDNIVLRDR 118
DB 11 DFRNDMPLS-----MELNLNMYRYTGTVMRFD--GERLTFNGDVEDLSAREP 58

QY 119 SNWVHGYPEIFYGNKPNANYATDGPILPSPKVSNLTDFTYLTISYKLEPKNGLPINFAIE 178
DB 59 ERYILGYPEFYGYKPEW-RHAAEG-TKLPLVSVSVESFTVELSFEIDHMPSLPLNFAME 116

QY 179 SWLTREAWRTTGINSDEQEWMIWYDGLQAGSKVKEIVPIVNGTPVYNATFEVWKAN 238
DB 117 TWLTREKYQVEASIGD-VEIMVWFYFNELTGPGKKGVEYTVSFELNGEKGKGIWELWHA 175

QY 239 IGWEYVAFRIKTIKEGTVPYGFATISVAANISS-----LPNYTELYLEDVEIGTEFGTP 294
DB 176 WNWDFLAFRLKPNVRKGRVRFNVKDFLDVAGEYLSRSTRVKDFDLYFTVWEIGTEFGSP 235

QY 295 STTSAHLEWITNITL 310
DB 236 ETKSARFGWTFNFSI 251

RESULT 10
QY97VS7 PRELIMINARY; PRT; 322 AA.
AC Q97VS7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulase (Endo 1,4 beta glucanase), putative (celB
DE (EC 3.2.1.4).
GN CELB OR CELS OR SSO2534.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
CC Sulfolobus.
CX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic V., Allard G.,
RA Aweez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Brauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49255 / DSM 5833 / MT-4;
RX Limauro D., Fiorentino G., Cannio R., Rossi M., Bartolucci S.;
RT "Identification and molecular characterization of thermophilic
RT endoglucanase gene, celS, from Sulfolobus solfataricus.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS006850; AAK42663.1; -.
DR EMBL; AJ296029; CAC39483.1; -.
```


QY 278 TELYLEDEVEIGTE 290
Db 230 PEMYLHAVETGFE 242

RESULT 13
Q54331 PRELIMINARY; PRT; 381 AA.
AC Q54331, 1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulase B precursor.
GN CELB
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=66;
RX MEDLINE=94288649; PubMed=9017952;
RA Wittmann S., Sharek F., Kluepfel D., Morosoli R.;
RT "Purification and characterization of the CelB endoglucanase from
Streptomyces lividans 66 and DNA sequence of the encoding gene.";
RL Appl. Environ. Microbiol. 60:1701-1703(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=66;
RA Sharek F.;
RL EMBL; U04629; AAB71950.1; -;
DR EMBL; PDB; 1NLR; 25-NOV-98.
DR PDB; 2NLR; 07-FEB-00.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR008965; ConA like lec gl.
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
DR SMART; SMO0637; CBD_II; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 381 AA; 39239 MW; AYE99BF590FA24EC CRC64;

Query Match 6.7%; Score 114.5; DB 2; Length 381;
Best Local Similarity 27.1%; Pred. No. 0.58;
Matches 52; Conservative 21; Mismatches 80; Indels 39; Gaps 10;

QY 125 YPEIPGNKPNVATDGPILPSKVSNTDFLTISYKLEPKNGLPINFAIESWLTRE 184
Db 97 YPSVFNGCHYTNCSFGTD----LPVRLDTVAAPSSISYGF--VDGAVYNASYDIWLDPT 150
QY 185 AWRTTGINSDEQVNMWIIYDQ--LQPSGSKYKEIVVPIIIVNGTPV--NATPEVW----- 235
Db 151 A-RTDGVN--QTEININWNRVGPPIQIGSPV-----GTASVGRTWVWSGGNS 197
QY 236 -----KANIGHEY-VAPRIKTIKEGVTIPIYGAFISVAANISLNPYTELYEDV 285
Db 198 NDVLSFVAFSAISGWSFDVDFVATVARGLAENDW-YLTSVQAGFPQNGAGLVNSF 256
QY 286 EIGTEFGTPST 297
Db 257 SSTVETGPGGT 268

RESULT 14

Q82K36 PRELIMINARY; PRT; 269 AA.
AC Q82K36;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative endo-1,4-beta-glucanase.
GN CELA4 OR SAV2568.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing Streptomyces
RT metabolites";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AF005031; BAC70279.1; -;
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR008965; ConA like lec gl.
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
KW Complete proteome.
SQ SEQUENCE 269 AA; 29362 MW; 07B90B00F577BC5F CRC64;

Query Match 6.6%; Score 111.5; DB 16; Length 269;
Best Local Similarity 22.0%; Pred. No. 0.64;
Matches 51; Conservative 38; Mismatches 96; Indels 47; Gaps 10;

QY 76 EPIEINLNLNATGAETMYNTLTSGVLHYVQQLDNIVLRDRSNWV-----HGYPEIFY 130
Db 50 KYIVNNLWQDKATGTQCWDDSRSG-----STISWGTDSWANNAGHDYDVKSY 101
QY 131 GNK-----PWANAYATDGPILPSKVSNTDFLTISYKLEPKNGLPINFAIESWLTRE 184
Db 102 ASSVLGWHGWKADKAATG---LPVRVGRKSVRTTWDFSVGAGPG-TWNVAYDLWLHAK 157
QY 185 AWRTTGINSDEQ-----EVMWMT-VYDGLQPSGSKYKEIVVPIIIVNGTPVNFATFEVKKAN 238
Db 158 N-----NADWQDQPTDEIMWLNROGGAGPLGTYGVSLSG-----GAMWDIYQGD 203
QY 239 IGHEYVAFRIKTIKEGVTIPIYGAFISVAANISLNPYTELYEDVEIGTE 290
Db 204 IGKVIYSFVRTTNTTKATLDLDDFTQALVRRKLLS---DDKYVSGIEAGTE 251

RESULT 15
C74705 PRELIMINARY; PRT; 239 AA.
AC C74705;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (CELLULASE)
DE (Endo-1,4-beta-glucanase A) (CARBOXYMETHYLCELLULASE).
GN EGLA OR CEKA.

